GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 18, 2002, 22:33:40 ; Search time 52 Seconds
(without alignments)
2901.634 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agctccgagcgtcattctcc.....agccaattttggngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
	1	57.4	11.7	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	54.4	11.1	320	4	US-09-165-264-7	Sequence 7, Appli
c	3	54.4	11.1	320	4	US-09-165-264-11	Sequence 11, Appl
c	4	53.8	10.9	318	4	US-09-165-264-12	Sequence 12, Appl
c	5	53.8	10.9	2721	6	5215881-2	Patent No. 5215881
c	6	53.8	10.9	8438	1	US-07-945-283-1	Sequence 1, Appli
c	7	53.6	10.9	320	4	US-09-165-264-13	Sequence 13, Appl
c	8	53.4	10.9	319	4	US-09-165-264-8	Sequence 8, Appli
	9	52.8	10.7	12001	1	US-08-458-568A-11	Sequence 11, Appl
c	10	52.6	10.7	320	4	US-09-165-264-14	Sequence 14, Appl
c	11	49.4	10.0	530	3	US-08-758-662-4	Sequence 4, Appli
c	12	49.4	10.0	4524	2	US-08-845-998-7	Sequence 7, Appli
c	13	49.4	10.0	4524	3	US-09-206-537-7	Sequence 7, Appli
c	14	49.4	10.0	4524	4	US-09-430-854-7	Sequence 7, Appli
	15	48	9.8	53526	3	US-08-658-136-2	Sequence 2, Appli
	16	48	9.8	53577	3	US-08-658-136-1	Sequence 1, Appli
	17	47.6	9.7	152331	3	US-09-128-155-16	Sequence 16, Appl
	18	47.4	9.6	2888	4	US-08-765-907A-1	Sequence 1, Appli
	19	47.2	9.6	530	3	US-08-758-662-4	Sequence 4, Appli
c	20	47.2	9.6	2823	1	US-08-398-008A-1	Sequence 1, Appli
c	21	47.2	9.6	2823	2	US-08-893-333-1	Sequence 1, Appli
	22	47.2	9.6	4524	2	US-08-845-998-7	Sequence 7, Appli
	23	47.2	9.6	4524	3	US-09-206-537-7	Sequence 7, Appli
	24	47.2	9.6	4524	4	US-09-430-854-7	Sequence 7, Appli
c	25	46.8	9.5	1926	4	US-09-249-585A-2	Sequence 2, Appli
c	26	46.8	9.5	2580	3	US-09-050-863-2	Sequence 2, Appli
c	27	46.8	9.5	2580	4	US-09-359-081-2	Sequence 2, Appli
	28	46.8	9.5	5452	2	US-09-130-114-1	Sequence 1, Appli
c	29	46.8	9.5	9600	4	US-08-910-647-1	Sequence 1, Appli
c	30	46.8	9.5	9600	4	US-09-620-925-1	Sequence 1, Appli
c	31	46.8	9.5	10596	1	US-07-884-811-15	Sequence 15, Appl
c	32	46.8	9.5	10596	1	US-07-885-971-15	Sequence 15, Appl
c	33	46.8	9.5	10596	1	US-08-087-783A-15	Sequence 15, Appl
c	34	46.8	9.5	10596	1	US-08-194-088B-15	Sequence 15, Appl
c	35	46.8	9.5	10596	2	US-08-194-087-15	Sequence 15, Appl
c	36	46.8	9.5	10596	5	PCT-US93-04648-15	Sequence 15, Appl
	37	46.4	9.4	1515	4	US-09-292-768-5	Sequence 5, Appli
c	38	45.6	9.3	289	4	US-09-007-005-17	Sequence 17, Appl
c	39	45.6	9.3	289	4	US-09-244-796-17	Sequence 17, Appl
	40	45.6	9.3	1028	4	US-08-118-200-1	Sequence 1, Appli
	41	45.6	9.3	1028	4	US-08-458-745-1	Sequence 1, Appli
c	42	45.6	9.3	2830	1	US-07-882-292-1	Sequence 1, Appli
c	43	45.6	9.3	2830	2	US-08-331-644-1	Sequence 1, Appli
c	44	45.6	9.3	2830	5	PCT-US93-04102-1	Sequence 1, Appli
c	45	45.4	9.2	2647	5	PCT-US93-06251-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 11.7%; Score 57.4; DB 1; Length 7218;

Best Local Similarity 2.1%; Pred. No. 0.00026;
Matches 4; Conservative 140; Mismatches 51; Indels 0; Gaps 0;

```
Qy       4 TCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTC 63
           | ||| : :: :::::: :::: : ::: :::: ::::: : :: :: :: ::
Db    1061 TGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1120

Qy      64 CTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCT 123
           :::::: : :::::::::::::::::: ::: ::: :: :::: ::::::::: :::
Db    1121 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1180

Qy     124 CGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGT 183
           : ::::: :  :::: ::::: ::::: ::::::: ::::: : :: :::: ::: :
Db    1181 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1240

Qy     184 GTCGTCCGCGGTCTT 198
           :: ::: :  ::::
Db    1241 YYYYYYYYYYYYYYY 1255
```

RESULT 2
US-09-165-264-7/c
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

Query Match 11.1%; Score 54.4; DB 4; Length 320;
Best Local Similarity 56.8%; Pred. No. 0.00065;
Matches 100; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
Qy       6 CGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCT 65
           | ||| ||   | || ||||  |       ||   | |||  |  || || ||  | ||
Db     307 CTAGCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 248

Qy      66 CCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCG 125
           ||||    | | || |  || || || ||| | |  || || | |||  ||||  || |
Db     247 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 188

Qy     126 CCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCC 181
           || || |   |||   ||||| | | | || || || || || | || || | |||
Db     187 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 132
```

RESULT 3
US-09-165-264-11/c
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

Query Match 11.1%; Score 54.4; DB 4; Length 320;
Best Local Similarity 59.9%; Pred. No. 0.00065;
Matches 91; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy      30 TTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCT 89
           || |||   ||| |||  |  || || ||  | || ||||    | | || |  || ||
Db     312 TTCAAAATCTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 253

Qy      90 CCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCT 149
           || ||| | |  || || | |||  ||||  || | || || |   |||   ||||| |
Db     252 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 193

Qy     150 CTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCC 181
           | | || || || || || | || || | |||
Db     192 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 161
```

RESULT 4
US-09-165-264-12/c
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

Query Match 10.9%; Score 53.8; DB 4; Length 318;
Best Local Similarity 60.6%; Pred. No. 0.00088;
Matches 88; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```
Qy      37 CTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCA 96
           ||| | | |  ||||| || ||  | || ||||    | | || |  || || || |||
Db     314 CTTACACACACCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 255

Qy      97 CTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCT 156
           | |  || || | |||  ||||  || | || || |   |||   ||||| | | | ||
Db     254 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 195

Qy     157 CCTCCGCCTCCGCGCCGCCTCGCCC 181
           || || || || | || || | |||
Db     194 CCCCCCCCCCCCCCCCCCCCCCCCC 170
```

RESULT 5
5215881-2/c
;Patent No. 5215881
; APPLICANT: CHEUNG, ANDREW K.
; TITLE OF INVENTION: PSEUDORABIES DIAGNOSIS PROBES
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/537,855
; FILING DATE: 13-JUN-1990
; FILING DATE:
;SEQ ID NO:2:
; LENGTH: 2721
5215881-2

Query Match 10.9%; Score 53.8; DB 6; Length 2721;

```
  Best Local Similarity   51.2%;  Pred. No. 0.0014;
  Matches  150;  Conservative    0;  Mismatches  142;  Indels    1;  Gaps    1;

Qy      52 CGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCC 111
           | || ||  |  |||| | | || || ||| |||  || | | | |   | |  |   ||
Db    1022 CCCCTCCACCGGCTCCTCGAGGCCCTTCTCCTCCGTCTTCTCTCCCCTCGTCCTCGGCCC 963

Qy     112 CTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGC 171
             ||| || ||||| |  ||    |     || || |    || ||| | ||  | ||
Db     962 GGTCCTCGTCCTCGTCCTCGTCCCCGTCCTCCTCCTCCTCTGCGTCCGCGGCGGCGGCCT 903

Qy     172 CGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGT 231
           ||   | | ||   || | ||| |      | ||        |   |||   |   |
Db     902 CGGTCCCCTCGGCCCGGCGGCGCTTGCCTCCCCGGCGCCTGCCTCCCCGGCCCGGTGGCC 843

Qy     232 TAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGC 291
             |||| | | |  ||    |  |||  ||||   |   | | | || ||    |||
Db     842 CTCCTCCTCCTCCTCTCCT-TCTCCTCCGCGCGGATCCCCCGGCCGGAGGTGGCTGCGGC 784

Qy     292 GACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGG 344
           | ||| |  | ||| || ||||| | ||  |||||||| || || | |||||
Db     783 GGCGGCGGAGGTGGCGGCGGTGGTGGAAGCGGCGGCGGCGGCCGCGGCGGAGG 731


RESULT 6
US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
;  GENERAL INFORMATION:
;    APPLICANT:  Cheung, Andrew K.
;    APPLICANT:  Wesley, Ronald D.
;    TITLE OF INVENTION:  Pseudorabies Virus Deletion Mutants
;    TITLE OF INVENTION:  Involving The EP0 and LLT Genes
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis P. Ribando
;      STREET:  1815 No. 5352596th University Street
;      CITY:  Peoria
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  61604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/945,283
;      FILING DATE:  19920911
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ribando, Curtis P
;      REGISTRATION NUMBER:  27976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  309-685-4011 ext.513
;      TELEFAX:  309-685-4128
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8438 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudorabies virus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  622..6495
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1099, "g")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1267, "t")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1381, "c")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1566, "c")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(7010, "g")
US-07-945-283-1

  Query Match             10.9%;  Score 53.8;  DB 1;  Length 8438;
  Best Local Similarity   51.2%;  Pred. No. 0.0017;
  Matches  150;  Conservative    0;  Mismatches  142;  Indels    1;  Gaps    1;

Qy      52 CGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCC 111
           | || ||  |  |||| | | || || ||| |||  || | | | |   | |  |   ||
Db    2365 CCCCTCCACCGGCTCCTCGAGGCCCTTCTCCTCCGTCTTCTCTCCCCTCGTCCTCGGCCC 2306

Qy     112 CTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGC 171
             ||| || ||||| |  ||    |     || || |    || ||| | ||  | ||
Db    2305 GGTCCTCGTCCTCGTCCTCGTCCCCGTCCTCCTCCTCCTCTGCGTCCGCGGCGGCGGCCT 2246

Qy     172 CGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGT 231
           ||   | | ||   || | ||| |      | ||        |   |||   |   |
Db    2245 CGGTCCCCTCGGCCCGGCGGCGCTTGCCTCCCCGGCGCCTGCCTCCCCGGCCCGGTGGCC 2186

Qy     232 TAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGC 291
             |||| | | |  ||    |  |||  ||||   |   | | | || ||    |||
Db    2185 CTCCTCCTCCTCCTCTCCT-TCTCCTCCGCGCGGATCCCCCGGCCGGAGGTGGCTGCGGC 2127

Qy     292 GACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGG 344
           | ||| |  | ||| || ||||| | ||  |||||||| || || | |||||
Db    2126 GGCGGCGGAGGTGGCGGCGGTGGTGGAAGCGGCGGCGGCGGCCGCGGCGGAGG 2074


RESULT 7
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-13

  Query Match             10.9%;  Score 53.6;  DB 4;  Length 320;
  Best Local Similarity   57.0%;  Pred. No. 0.00098;
  Matches   98;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy      10 CGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCC 69
           ||||    ||| ||||  |       ||   | |||  |  || || ||  | || ||||
Db     309 CGTCCACGTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 250

Qy      70 AATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTC 129
             | | || |  || || || ||| | |  || || | |||  ||||  || | || |
Db     249 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 190

Qy     130 CGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCC 181
           | |   |||   ||||| | | | || || || || || | || || | |||
```

```
Db    189 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 138


RESULT 8
US-09-165-264-8/c
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 319
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
S-09-165-264-8

  Query Match             10.9%;  Score 53.4;  DB 4;  Length 319;
  Best Local Similarity   60.8%;  Pred. No. 0.0011
  Matches   87;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      39 TGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACT 98
           |||| | |||  || || ||  | || ||||    | | || |  || || || ||| |
Db     309 TGCTGCAGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 250

Qy      99 CGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCC 158
           |  || || | |||  |||| || | || || |   |||  ||||| | | | || ||
Db     249 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 190

Qy     159 TCCGCCTCCGCGCCGCCTCGCCC 181
           || || || | || || | |||
Db     189 CCCCCCCCCCCCCCCCCCCCCCC 167


RESULT 9
US-08-458-568A-11
; Sequence 11, Application US/08458568A
; Patent No. 5821339
;  GENERAL INFORMATION:
;    APPLICANT:  Schaffer, Priscilla A.
;    APPLICANT:  Yeh, Lily
;    TITLE OF INVENTION:  Compositions and Methods for Treatment of Herpesvirus
     TITLE OF INVENTION:  Infections
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,568A
;      FILING DATE:  02-JUNE-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/065,146
;      FILING DATE:  05-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leary Ph.D., Kathryn R.
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  DFCI-0029
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpes simplex virus
;      STRAIN:  Herpes Simplex Virus Type 1
US-08-458-568A-11

  Query Match             10.7%;  Score 52.8;  DB 1;  Length 12001;
  Best Local Similarity   52.8%;  Pred. No. 0.003;
  Matches  114;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy      52 CGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCC 111
           | || ||  |  |||| |    | ||||||  ||  |||| ||| |    || || | |
Db    1347 CTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCG 1406

Qy     112 CTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGC 171
           || | ||  || | |||||  |   | |   |||||||||   || || |  || | || |
Db    1407 CTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCCGCGGCCC 1466

Qy     172 CGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGT 231
           |||| | | ||       ||||  |    ||       |  ||||  |||      |  |
Db    1467 CGCCCCCCACGCCCGCCGCGCGCGCGCACGCCGCCCGGACCGCCGCCCGCCTTTTTTGCG 1526

Qy     232 TAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAG 267
             |  | ||  ||| |  ||  ||| | |||  ||
Db    1527 CGCGCGCGCGCCCGCGGGGGGCCCGGGCTGCCACAG 1562


RESULT 10
US-09-165-264-14/c
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
;  APPLICANT:  Vinayagamoorthy, Thuraiayah
;  TITLE OF INVENTION:  Multi-Loci Genomic Analysis
;  FILE REFERENCE:  44747
;  CURRENT APPLICATION NUMBER: US/09/165,264
;  CURRENT FILING DATE:  1998-10-01
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 320
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14

  Query Match             10.7%;  Score 52.6;  DB 4;  Length 320;
  Best Local Similarity   56.7%;  Pred. No. 0.0016;
  Matches   97;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy      11 GTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCA 70
           | ||  | || ||||  |        ||   | |||  |  || || ||  | || ||||
Db     306 GGCACACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 247

Qy      71 ATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCC 130
           | | || |  || || || ||| | |  || || | |||  ||||  || | || ||
Db     246 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 187

Qy     131 GCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCC 181
             |   |||   ||||| | | | || || || || || | || || | |||
```

```
Db    186 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 136


RESULT 11
US-08-758-662-4/c
; Sequence 4, Application US/08758662
; Patent No. 6114150
;  GENERAL INFORMATION:
;    APPLICANT:  Weissman, Sherman
;    APPLICANT:  Baskaran, Namadev
;    TITLE OF INVENTION:  AMPLIFICATION OF NUCLEIC ACIDS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  701 Fifth Avenue, 6300 Columbia Center
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/758,662
;      FILING DATE:  29-NOV-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6114150tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  390036.402C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  530 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-758-662-4

  Query Match             10.0%;  Score 49.4;  DB 3;  Length 530;
  Best Local Similarity   49.4%;  Pred. No. 0.0093;
  Matches  128;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy      44 CCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCC 103
           ||  |  ||   |||| | || || |    |   || |  || || || ||| |  ||||
Db     460 CCCGCCGCCCGGGCCGCCGCCGCCGCCCCCCGGGCCGCCACCGCCGCCTCCCCCGGGGCC 401

Qy     104 ACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGC 163
           || |  ||    ||| |  |||| ||    | | | | ||| |     | || |  ||
Db     400 GCCGCCGCCCGGGCCGCCGCCGCCGCCCGGGCCGCCACCGCCCCCCGGGCCGTCGTGGGC 341

Qy     164 CTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATG 223
           |  |||||||  | |    |  ||| | |     ||    || |||| |||    |
Db     340 GCCGCCGCCGCCGGCCGCCGCGCCCGCGCCGCCCGCGCCGGCCATGAGCGCGAGCGACGG 281

Qy     224 AGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGG 283
           |    | |  || |  |     ||| |||   |   ||| |   |  | ||  |||| |
Db     280 CGAGGAGATGTGGTCCAGCAGATCGCCGGGTTCGAGCGCCTGGTGGTGGTGGTGGTGGTG 221

Qy     284 GAGTTGGCGACGGAGCGGC 302
           | | ||| |    || |||
Db     220 GTGGTGGTGCGCCAGAGGC 202


RESULT 12
US-08-845-998-7/c
; Sequence 7, Application US/08845998
; Patent No. 5879892
;  GENERAL INFORMATION:
;    APPLICANT:  Van Baren, Nicolas
;    APPLICANT:  Coulie, Pierre G.
;    APPLICANT:  De Smet, Charles
;    APPLICANT:  Lucas, Sophie
;    APPLICANT:  Boon, Thierry
;    TITLE OF INVENTION:  LEUKEMIA ASSOCIATED GENES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/845,998
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Van Amsterdam, John R.
;      REGISTRATION NUMBER:  40,212
;      REFERENCE/DOCKET NUMBER:  L0461/7008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)720-3500
;      TELEFAX:  (617)720-2441
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4524 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  174..1433
US-08-845-998-7

  Query Match             10.0%;  Score 49.4;  DB 2;  Length 4524;
  Best Local Similarity   49.4%;  Pred. No. 0.014;
  Matches  128;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy      44 CCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCC 103
           ||  |  ||   |||| | || || |    |   || |  || || || ||| |  ||||
Db     719 CCCGCCGCCCGGGCCGCCGCCGCCGCCCCCCGGGCCGCCACCGCCGCCTCCCCCGGGGCC 660

Qy     104 ACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGC 163
           || |  ||    ||| |  |||| ||    | | | | ||| |     | || |  ||
Db     659 GCCGCCGCCCGGGCCGCCGCCGCCGCCCGGGCCGCCACCGCCCCCCGGGCCGTCGTGGGC 600

Qy     164 CTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATG 223
           |  |||||||  | |    |  ||| | |     ||    || |||| |||    |
Db     599 GCCGCCGCCGCCGGCCGCCGCGCCCGCGCCGCCCGCGCCGGCCATGAGCGCGAGCGACGG 540

Qy     224 AGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGG 283
           |    | |  || |  |     ||| |||   |   ||| |   |  | ||  |||| |
Db     539 CGAGGAGATGTGGTCCAGCAGATCGCCGGGTTCGAGCGCCTGGTGGTGGTGGTGGTGGTG 480

Qy     284 GAGTTGGCGACGGAGCGGC 302
           | | ||| |    || |||
Db     479 GTGGTGGTGCGCCAGAGGC 461


RESULT 13
US-09-206-537-7/c
; Sequence 7, Application US/09206537
; Patent No. 6130052
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Van Baren, Nicolas
;    APPLICANT:  Coulie, Pierre G.
;    APPLICANT:  De Smet, Charles
;    APPLICANT:  Lucas, Sophie
;    APPLICANT:  Boon, Thierry
;    TITLE OF INVENTION:  LEUKEMIA ASSOCIATED GENES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/206,537
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/845,998
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Van Amsterdam, John R.
;      REGISTRATION NUMBER:  40,212
;      REFERENCE/DOCKET NUMBER:  L0461/7008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)720-3500
;      TELEFAX:  (617)720-2441
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4524 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  174..1433
US-09-206-537-7

  Query Match             10.0%;  Score 49.4;  DB 3;  Length 4524;
  Best Local Similarity   49.4%;  Pred. No. 0.014;
  Matches  128;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy      44 CCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCC 103
           ||  |  ||   |||| | || || |    |   || |  || || || ||| |  ||||
Db     719 CCCGCCGCCCGGGCCGCCGCCGCCGCCCCCGGGCCGCCACCGCCGCCTCCCCCGGGGCC 660

Qy     104 ACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGC 163
           || |  ||   ||| |  |||| ||     | | | | ||| |     | || |  ||
Db     659 GCCGCCGCCCGGGCCGCCGCCGCCGCCCGGGCCGCCACCGCCCCCCGGGCCGTCGTGGGC 600

Qy     164 CTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATG 223
             |  |||||||   | |    |  ||| | |     ||    || |||| |||    |
Db     599 GCCGCCGCCGCCGGCCGCCGCGCCCGCGCCGCCCGCGCCGGCCATGAGCGCGAGCGACGG 540

Qy     224 AGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGG 283
           |    | |  || |  |     ||| |||   |   ||| |   |  | ||  |||| |
Db     539 CGAGGAGATGTGGTCCAGCAGATCGCCGGGTTCGAGCGCCTGGTGGTGGTGGTGGTGGTG 480

Qy     284 GAGTTGGCGACGGAGCGGC 302
           | | ||| |    || |||
Db     479 GTGGTGGTGCGCCAGAGGC 461
```

```
RESULT 14
US-09-430-854-7/c
; Sequence 7, Application US/09430854
; Patent No. 6271019
;  GENERAL INFORMATION:
;    APPLICANT:  Van Baren, Nicolas
;    APPLICANT:  Coulie, Pierre G.
;    APPLICANT:  De Smet, Charles
;    APPLICANT:  Lucas, Sophie
;    APPLICANT:  Boon, Thierry
;    TITLE OF INVENTION:  LEUKEMIA ASSOCIATED GENES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/430,854
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/845,998
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Van Amsterdam, John R.
;      REGISTRATION NUMBER:  40,212
;      REFERENCE/DOCKET NUMBER:  L0461/7008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)720-3500
;      TELEFAX:  (617)720-2441
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4524 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  174..1433
US-09-430-854-7

  Query Match             10.0%;  Score 49.4;  DB 4;  Length 4524;
  Best Local Similarity   49.4%;  Pred. No. 0.014;
  Matches  128;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy      44 CCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCC 103
           ||  |  ||   |||| | || || |    |   || |  || || || ||| |  ||||
Db     719 CCCGCCGCCCGGGCCGCCGCCGCCGCCCCCGGGCCGCCACCGCCGCCTCCCCCGGGGCC 660

Qy     104 ACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGC 163
           || |  ||   ||| |  |||| ||     | | | | ||| |     | || |  ||
Db     659 GCCGCCGCCCGGGCCGCCGCCGCCGCCCGGGCCGCCACCGCCCCCCGGGCCGTCGTGGGC 600

Qy     164 CTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATG 223
             |  |||||||   | |    |  ||| | |     ||    || |||| |||    |
Db     599 GCCGCCGCCGCCGGCCGCCGCGCCCGCGCCGCCCGCGCCGGCCATGAGCGCGAGCGACGG 540

Qy     224 AGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGG 283
           |    | |  || |  |     ||| |||   |   ||| |   |  | ||  |||| |
Db     539 CGAGGAGATGTGGTCCAGCAGATCGCCGGGTTCGAGCGCCTGGTGGTGGTGGTGGTGGTG 480

Qy     284 GAGTTGGCGACGGAGCGGC 302
```

```
          | | ||| |    || |||
Db    479 GTGGTGGTGCGCCAGAGGC 461


RESULT 15
US-08-658-136-2
; Sequence 2, Application US/08658136
; Patent No. 6071717
;  GENERAL INFORMATION:
;    APPLICANT:  KLINGER, KATHERINE W
;    APPLICANT:  LANDES, GREGORY M
;    APPLICANT:  BURN, TIMOTHY C
;    APPLICANT:  CONNORS, TIMOTHY D
;    APPLICANT:  DACKOWSKI, WILLIAM
;    APPLICANT:  GERMINO, GREGORY
;    APPLICANT:  QIAN, FENG
;    TITLE OF INVENTION:  POLYCYSTIC KIDNEY DISEASE GENE
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENZYME CORPORATION
;      STREET:  ONE MOUNTAIN ROAD
;      CITY:  FRAMINGHAM
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  01701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/658,136
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LASSEN, ELIZABETH
;      REGISTRATION NUMBER:  31,845
;      REFERENCE/DOCKET NUMBER:  GEN4-17.8
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  508-872-8400
;      TELEFAX:  508-872-5415
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  53526 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
 S-08-658-136-2

  Query Match             9.8%;  Score 48;  DB 3;  Length 53526;
  Best Local Similarity   53.8%;  Pred. No. 0.048;
  Matches   99;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
           ||||   | || | |  |  |||  | |     |    |  ||  | | |  || ||  |
Db   34585 CTCCTCCCCTCCTCCCTCCTCCCCTCCTCCCCCCTCCTCCTCCCCCTCCTCCCTCCTCCC 34644

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
            |  |||  | |||  |    ||||||||  || | |  || || | ||| ||| |  ||
Db   34645 TCCTCCCCCTCCTCCTCCTCCCCTCCTCCCTCCTCCCCTCCTCCCCTCCCCTCCTCCCCC 34704

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182
           || || |  |   ||    || || | ||| |||||||| ||||  | || ||   |||
Db   34705 TCCCCCCTCCCTTCCTCCCCCTCCCCCCTCCCCTCCTCCCCCTCTCCTCCTCCCATCCCT 34764

Qy     183 TGTC 186
             ||
Db   34765 CCTC 34768


Search completed: November 19, 2002, 00:27:01
Job time : 116 secs
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2002, 17:14:44 ; Search time 13.41 Seconds
(without alignments)
975.929 Million cell updates/sec

Title: US-09-686-522A-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	334	19.3	261	1	YHEB_CHLVI	P56160	chlorobium
2	292	16.9	271	1	SUHB_PSEAE	Q9hxi4	pseudomonas
3	287	16.6	267	1	SUHB_PASMU	Q9cnv8	pasteurella
4	279	16.1	264	1	SUHB_AQUAE	O67791	aquifex aeo
5	272	15.7	267	1	SUHB_HAEIN	P44333	haemophilus
6	265.5	15.3	268	1	MYO3_LYCES	P54928	lycopersico
7	262	15.1	267	1	SUHB_SALTY	P58537	salmonella
8	258	14.9	267	1	SUHB_ECOLI	P22783	escherichia
9	248	14.3	267	1	SUHB_VIBCH	Q9kty5	vibrio chol
10	247.5	14.3	273	1	MYO1_LYCES	P54926	lycopersico
11	238	13.8	265	1	MYO2_LYCES	P54927	lycopersico
12	235	13.6	270	1	MYOP_MESCR	O49071	mesembryant
13	234	13.5	275	1	SUHB_XYLFA	Q9pam0	xylella fas
14	229.5	13.3	256	1	SUHB_THEMA	O33832	thermotoga
15	228	13.2	266	1	SUHB_RHILO	Q98f59	rhizobium l
16	221.5	12.8	265	1	SUHB_BACSU	Q45499	bacillus su
17	218	12.6	259	1	STRO_STRGR	P29785	streptomyce
18	212.5	12.3	287	1	SUHB_SYNY3	P74158	synechocyst
19	211	12.2	266	1	SUHB_RHIME	Q92m71	rhizobium m
20	211	12.2	277	1	MYOP_BOVIN	P20456	bos taurus
21	209	12.1	285	1	MYOP_XENLA	P29219	xenopus lae
22	206	11.9	246	1	CYSQ_ECOLI	P22255	escherichia
23	203.5	11.8	277	1	MYOP_MOUSE	O55023	mus musculu
24	202.5	11.7	261	1	SUHB_NEIMB	Q9jz07	neisseria m
25	201.5	11.6	261	1	SUHB_NEIMA	Q9ju03	neisseria m
26	200	11.6	252	1	SUHB_METJA	Q57573	methanococc
27	196.5	11.4	288	1	MYO2_HUMAN	O14732	homo sapien
28	193.5	11.2	277	1	MYOP_HUMAN	P29218	homo sapien
29	193.5	11.2	277	1	MYOP_RAT	P97697	rattus norv
30	190.5	11.0	269	1	CYSQ_ACTAC	P70714	actinobacil
31	190	11.0	246	1	CYSQ_SALTY	P26264	salmonella
32	183.5	10.6	256	1	SUHB_CAUCR	Q9a3d5	caulobacter
33	181.5	10.5	341	1	YSNM_CAEEL	Q19420	caenorhabdi
34	181	10.5	275	1	Y4FL_RHISN	P55450	rhizobium s
35	180.5	10.4	340	1	QAX_NEUCR	P11634	neurospora
36	175	10.1	280	1	SUHB_METTH	O26957	methanobact
37	171	9.9	353	1	DPNP_ARATH	Q42546	arabidopsis
38	164.5	9.5	252	1	SUHB_ARCFU	O30298	archaeoglob
39	161	9.3	266	1	SUHB_BUCAI	P57372	buchnera ap
40	158.5	9.2	358	1	DPNP_ORYSA	Q40639	oryza sativ
41	157	9.1	290	1	SUHB_MYCTU	O07203	mycobacteri
42	151.5	8.8	357	1	HAL2_YEAST	P32179	saccharomyc
43	144.5	8.4	265	1	CYSQ_BUCAI	P57624	buchnera ap
44	142	8.2	295	1	YHK6_YEAST	P38710	saccharomyc
45	141.5	8.2	330	1	QUTG_EMENI	P25416	emericella

ALIGNMENTS

RESULT 1
YHEB_CHLVI
ID YHEB_CHLVI STANDARD; PRT; 261 AA.
AC P56160;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 28.2 kDa protein in hemB 3'region.
OS Chlorobium vibrioforme.
OC Bacteria; Green sulfur bacteria; Chlorobium.
OX NCBI_TaxID=1098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F. THIOSULFATOPHILUM / NCIB 8327;
RX MEDLINE=96215215; PubMed=8626508;
RA Rhie G.-E., Avissar Y.J., Beale S.I.;
RT "Structure and expression of the Chlorobium vibrioforme hemB gene and
RT characterization of its encoded enzyme, porphobilinogen synthase.";
RL J. Biol. Chem. 271:8176-8182(1996).
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; U38348; -; NOT_ANNOTATED_CDS.
DR HSSP; P29218; 2HHM.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 2.
DR PRINTS; PR00378; INOSPHPHTASE.
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 261 AA; 28243 MW; D25E7D223024095F CRC64;

Query Match 19.3%; Score 334; DB 1; Length 261;
Best Local Similarity 53.9%; Pred. No. 5.8e-19;
Matches 78; Conservative 43; Mismatches 91; Indels 18; Gaps 4;

```
Qy      81 VAQRAADAAGEVLRKYF-RQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: ||::   || |: :::  | | :||| ||| ||| :   |   ||   :|||
Db       9 LALELAEKAGKLTLDYFGRRSLQVFSKRDDTPVTEADRNAEELIRQGISAKFPDDGLFGE 68

Qy     140 E-------NGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           |       || |         |::|||||:||| | ||:| :|||   |   :|||: |
Db      69 EFDEHPSGNGRR---------WIIDPIDGTRSFIHGVPLYGVMIALEVEGAMQLGVINFP 119

Qy     193 ILRERWIGVDGKQTTLNGQEISVRS-CNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKV 251
           | | :    |    :|| : | :      |   ::|   :| :  :     ::|   :
Db     120 ALGELYQAERGSGAFMNGSPVQVSAIAENSASTVVFTEKEYLLDPPSNHPVDQLRIDAGL 179
```

```
Qy     252 PLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
           ||| : |:|||  :: |:  : |:|  :::|::| |||   |:||
Db     180 VRGWGDCYGHMLVASGRAEVAVDKIMSPWDCAAVIPIVEEAGGCCFDYRG 229


RESULT   2
SUHB_PSEAE
ID   SUHB_PSEAE     STANDARD;      PRT;   271 AA.
AC   Q9HXI4;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR PA3818.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
     SEQUENCE FROM N.A.
     STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004799; AAG07205.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   271 AA;  29521 MW;  EE1D43CFFEB871EA CRC64;


  Query Match             16.9%;  Score 292;  DB 1;  Length 271;
  Best Local Similarity   29.7%;  Pred. No. 1.1e-15;
  Matches   82;  Conservative   54;  Mismatches  106;  Indels   34;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEII---DKEDHSPVTIADREAEEAMVSVILKSFP 132
           :: :| ||| :|||::   ||  :|:::|   :|:    ,|  || ||: :|: : |:::
Db       4 MLNIALRAARSAGELI---FRSIERLDVISVNEKDAKDYVTEVDRAAEQTIVAALRKAYP 60

Qy     133 THAIFGEENGW-RCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
           |||| ||| |:   :   ||::||:||:||| :|| | | |   ||  : |:    |:
Db      61 THAIMGEEGGFIEGSGEGADYLWVIDPLDGTTNFIHGVPHFAVSIACKYKGRLEHAVVLD 120

Qy     192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHL-FEADAEDAFIR-----V 245
           |: :| :    |:  ||| : |     |  | | |  |    : |  | ::      |
Db     121 PVRQEEFTASRGRGAALNGRRLRVSGRKSLEGALLGTGFPFRDNQIDNLDNYLNMFRSLV 180

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG---- 301
                :   |     | :|:|  |   | ||  :|  :   ::: ||| ::|: |
```

```
Db     181 GQTAGIRRAGAASLDLAYVAAGRYDAFWEFGLSEWDMAAGALLVQEAGGLVSDFTGSHEF 240

Qy     302 -DKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
            :| |              : ||: :  |  |  ::
Db     241 LEKGH-------------IVAGNTKCFKALLTTIQ 262


RESULT   3
SUHB_PASMU
ID   SUHB_PASMU     STANDARD;      PRT;   267 AA.
AC   Q9CNV8;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR PM0315.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006067; AAK02399.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29470 MW;  E7070FAE3BA589CA CRC64;


  Query Match             16.6%;  Score 287;  DB 1;  Length 267;
  Best Local Similarity   30.6%;  Pred. No. 2.7e-15;
  Matches   83;  Conservative   49;  Mismatches  111;  Indels   28;  Gaps    6;

Qy      78 LVEVAQRAADAAGEVLRKYFRQR--VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || |: | : :|  ::   |  :  ||  |: :|||:: || ||:| |
Db       4 MLNIAIRAARKAGNVIAKGYERRDDLQTTLKSTNDYVTNIDKASEEAIIEVIRKSYPDHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  ||:|    ::| |  ||:||:|||  :|: | | |   ||:  |:  :||:  ||
Db      64 IITEESGALEGKDS-DIQWVIDPLDGTTNFVKGLPHFSVSIAIRVKGRTEVGVVYDPIRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           | :  | |:   :|   : | :   ||   | |  |           |  : | ::
Db     123 ELFTAVRGEGAKINDMRLRVENKRDLAGTVLTTGFPFKQTRLMPMQFAMMNNLIQ----- 177

Qy     256 CDCYAY----------ALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
            ||  :           :|:| ||   | |:| :|  :   ::  ||| : |: |
Db     178 -DCADFRRMGSAALDLCYVAAGRVDGYFEVGVKAWDIAAGDLIVREAGGLVCDFNG---- 232

Qy     306 WPVTAESRPTSFNVVAAGDARVHKEALDALR 336
               |  || ::|||   |: || |: ::
```

```
Db    233 ----GHSYLTSGHLVAAA-PRIVKEILNKIQ 258


RESULT   4
SUHB_AQUAE
ID   SUHB_AQUAE     STANDARD;      PRT;   264 AA.
AC   O67791;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR IMP2 OR AQ_1983.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000766; AAC07753.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; FALSE_NEG.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   264 AA;  29336 MW;  10F32D1D3B81B76A CRC64;


  Query Match             16.1%;  Score 279;  DB 1;  Length 264;
  Best Local Similarity   28.5%;  Pred. No. 1.1e-14;
  Matches   75;  Conservative   54;  Mismatches  106;  Indels   28;  Gaps    7;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  |||| ||
Db       5 KKYLEVAKIAALAGGQVLKENFGKVKKENIEEKGEKDFVSYVDKTSEERIKEVILKFFPD 64

Qy     134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           | : ||| |   || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 HEVVGEEMG---AEGSGSEYRWFIDPLDGTKNYINGFPIFAVSVGLVKGEEPIVGAVYLP 121

Qy     193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRV 245
           : : |  |    :||: | |:    |  | :    |      |          :| |  |
Db     122 YFDKLYWGAKGLGAYVNGKRIKVKDNESLKHAGVVYGFPSRSRRDISIYLNIFKDVFYEV 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
           ::  |  |        ::| |  | ::|  :||:|  : : ::: |||
Db     182 -GSMRRP--GAAAVDLCMVAEGIFDGMMEFEMKPWDIAAGLVILKEAGG---------- 227

Qy     306 WPVTAESRPTSFNVVAAGDARVH 328
           |    |   : : ||:  :|
Db     228 -VYTLVGEPFGVSDIIAGNKALH 249


RESULT   5
SUHB_HAEIN
ID   SUHB_HAEIN     STANDARD;      PRT;   267 AA.
AC   P44333;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR HI0937.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32775; AAC22595.1; -.
DR   HSSP; P29218; 2HHM.
DR   TIGR; HI0937; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00378; INOSPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29499 MW;  DF974E0DAB0CCEA6 CRC64;


  Query Match             15.7%;  Score 272;  DB 1;  Length 267;
  Best Local Similarity   30.8%;  Pred. No. 3.9e-14;
  Matches   84;  Conservative   44;  Mismatches  113;  Indels   32;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFRQR--VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || |: | : :|  :|   |  :  ||  |: :|  :: || ||:| |
Db       4 MLNIAIRAARKAGNVIAKNYERRDAIESTQKGINDYVTNVDKASEAEIIEVIRKSYPDHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |       :|  |::||:|||::|:|| | |   ||:    :  :||:  ||
Db      64 IITEETG-AIEGKDSDVQWIIDPLDGTRNFMTGLPHFSVSIAVRVKNRTEVGVVYDPIRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPH----------LFEADAEDAF-I 243
           | : | |:   ||   : | |   :  : | |  |            :  |  |||
Db     123 ELFTAVRGEGAKLNEVRLRVDSKREIQGSILATGFPFKQPKLMPAQFAMMNALIEDAADF 182

Qy     244 RVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
```

```
Db    183 RRTGSAALDL----CY----VASNRIDGYFEMGLKAWDCAAGDLIVREAGGLVCDFD--- 231

Qy    304 LHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
Db    232 -----AGNSYLRSGNIIAA-PSRVIKEMLNKIR 258


RESULT   6
MYO3_LYCES
ID   MYO3_LYCES     STANDARD;      PRT;   268 AA.
AC   P54928;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase 3 (EC 3.1.3.25) (IMPase 3) (IMP
DE   3) (Inositol monophosphatase 3).
GN   IMP3.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
     Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
     Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. VFNT CHERRY;
RX   MEDLINE=96351935; PubMed=8718627;
RA   Gillaspy G.E., Keddie J.S., Oda K., Gruissem W.;
RT   "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RT   by a multigene family.";
RL   Plant Cell 7:2175-2185(1995).
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARI Y).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39059; AAB19031.1; -.
DR   HSSP; P29218; 2HHM.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00378; INOSPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Lithium; Multigene family.
SQ   SEQUENCE   268 AA;  29054 MW;  BCB635A029804901 CRC64;


  Query Match             15.3%;  Score 265.5;  DB 1;  Length 268;
  Best Local Similarity   30.2%;  Pred. No. 1.2e-13;
  Matches   83;  Conservative   48;  Mismatches  117;  Indels   27;  Gaps    8;

Qy     74 ATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
Db      6 SVEQFLDVAVEAAKKAGEIIREGFYKTKHVEHKGMVDLVTETDKACEDFIFNHLKQRFPS 65

Qy    134 HAIFGEENGWRCA--ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
Db     66 HKFIGEETTAACGNFELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIEKKPTVGVVYN 125

Qy    192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT---TSPHLFEADA-----EDAFI 243
```

```
Db    126 PIIDELFTGIDGKGAFLNGKPIKVSSQSELVKALLATEAGTNRDKLVVDATTGRINSLLF 185

Qy    244 RVRNKVKVPLYGCDCYAYAL--LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
Db    186 KVRS-----LRMCGSCALNLCGVACGRLDLFYELEFGGPWDVAGGAVIVKEAGGFVFDPS 240

Qy    301 GDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
Db    241 GSE--FDLTARR-------VAATNAHLKDAFIKAL 266


RESULT   7
SUHB_SALTY
ID   SUHB_SALTY     STANDARD;      PRT;   267 AA.
AC   P58537;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR STM2546.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008815; AAL21440.1; -.
DR   StyGene; SG?????; suhB.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29158 MW;  FC46D8A8298BB334 CRC64;


  Query Match             15.1%;  Score 262;  DB 1;  Length 267;
  Best Local Similarity   30.0%;  Pred. No. 2.3e-13;
  Matches   80;  Conservative   44;  Mismatches  123;  Indels   20;  Gaps    7;

Qy     78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
Db      4 MLTIAVRAARKAGNVIAKNYETPDAVEASQKGSNDFVTNVDKAAEAVIIDTIRKSYPQHT 63

Qy    136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
Db     64 IITEESGEHVGTDQ-DVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy    196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKVKVPL- 253
```

```
Db    123 ELFTATRGQGAQLNGYRLRGSTARDLDGTILATGFP--FKAKQYATTYINIIGKLFTECA 180

Qy    254 ----YGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
                |      | :|:| ||   | ||:|:|| :   ::  ||| ::|: |   :: :|
Db    181 DFRRTGSAALDLAYVAAGRVDGFFEIGLRPWDFAAGELLVREAGGIVSDFTGGH-NYMMT 239

Qy    310 AESRPTSFNVVAAGDARVHKEALDALR 336
                  : ||: || |  |  :|
Db    240 GN--------IVAGNPRVVKAMLANMR 258


RESULT   8
SUHB_ECOLI
ID   SUHB_ECOLI     STANDARD;      PRT;   267 AA.
AC   P22783; P77511;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR SSYA OR B2533.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=90202735; PubMed=2138605;
RA   Yano R., Nagai H., Shiba K., Yura T.;
RT   "A mutation that enhances synthesis of sigma 32 and suppresses
RT   temperature-sensitive growth of the rpoH15 mutant of Escherichia
RT   coli.";
RL   J. Bacteriol. 172:2124-2130(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=95095968; PubMed=8002619;
RA   Matsuhisa A., Suzuki N., Noda T., Shiba K.;
RT   "Inositol monophosphatase activity from the Escherichia coli suhB gene
RT   product.";
RL   J. Bacteriol. 177:200-205(1995).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=20213266; PubMed=10747806;
RA   Chen L., Roberts M.F.;
RT   "Overexpression, purification, and analysis of complementation
RT   behavior of E. coli SuhB protein: comparison with bacterial and
RT   archaeal inositol monophosphatases.";
RL   Biochemistry 39:4145-4153(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34828; AAA67506.1; -.
DR   EMBL; AE000339; AAC75586.1; -.
DR   EMBL; D90883; BAA16427.1; -.
DR   EMBL; D90884; BAA16435.1; -.
DR   PIR; A35158; A35158.
DR   HSSP; P29218; 2HHM.
DR   SWISS-2DPAGE; P22783; COLI.
DR   EcoGene; EG10983; suhB.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00378; INOSPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
FT   CONFLICT    141    141       R -> L (IN REF. 1).
SQ   SEQUENCE   267 AA;  29172 MW;  8FEC3508BD111301 CRC64;


  Query Match             14.9%;  Score 258;  DB 1;  Length 267;
  Best Local Similarity   29.6%;  Pred. No. 4.7e-18;
  Matches   79;  Conservative   44;  Mismatches   1;  Indels   20;  Gaps    7;

Qy      78 LVEVAQRAADAAGEVLRYFR--QRVELIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| ||| || :: | :      ||    |  :  ||  |: ||  ::  | ||:| |
Db       4 MLNIAVRAAKKAGNLIAKNYETPDAVEASQKGSNDFVTNVDKAAEAVIIDTIRKSYPQHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  ||:|         |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db      64 IITEESG-ELEGTDQDVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy     196 ERWIGVDGKQTTLNGQETSVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKV----- 249
           | :    |:  ||| :  :  |   | |  |  |:|      :| :  |:
Db     123 ELFTATRGQGAQLNGYRLRGSTARDLDGTILATGFP--FKAKQYATTYINIVGKLFNECA 180

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
                 |      | :|:| ||   | ||:|:|| :   ::  ||| ::|: |   :: :|
Db     181 DFRRTGSAALDLAYVAAGRVDGFFEIGLRPWDFAAGELLVREAGGIVSDFTGGH-NYMLT 239

Qy     310 AESRPTSFNVVAAGDARVHKEALDALR 336
                  : ||: || |  |  :|
Db     240 GN--------IVAGNPRVVKAMLANMR 258


RESULT   9
SUHB_VIBCH
ID   SUHB_VIBCH     STANDARD;      PRT;   267 AA.
AC   Q9KTY5;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   VC0745.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
```

```
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004160; AAF93910.1; ALT_INIT.
DR   HSSP; P29218; 1IMF.
DR   TIGR; VC0745; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29088 MW;  88AD8CDF78A7AD87 CRC64;


  Query Match             14.3%;  Score 248;  DB 1;  Length 267;
  Best Local Similarity   28.9%;  Pred. No. 2.8e-12;
  Matches   79;  Conservative   41;  Mismatches  121;  Indels   32;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  ||  : |     ::::   |  :  ||  |:|||  :|| |  |:| |
Db       4 MLNIAIRAARKAGNHIAKSLENAEKIQTTQKGSNDFVTNVDKEAEAIIVSTIKSSYPEHC 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |       :  |::||:||| :|: | | |   ||:   ||  :  :  |:
Db      64 IIAEEGG-LIEGKDKEVQWIIDPLDGTTNFVKGFPHFAVSIAVRFRGKTEVACVYDPMTN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           | :    |    ||   | |:    |  | | |  |   :  :| :|::: : : |
Db     123 ELFTAQRGAGAQLNNARIRVQPIKDLQGAVLATAFPFKQKQHSE-SFMKILSAMFVE--- 178

Qy     256 CD-----------CYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           |             ||    ||:  ||   | ||||:|  :   :   ||  :||: |
Db     179 CADFRRTGSAALDLCY----LAANRVDGYFELGLKPWDMAAGELIAREAGAIVTDFAG-- 232

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                   | |:||:    | |  |  :|
Db     233 ------GTDYMQSGNIVASSPRGV-KAILQHIR 258


RESULT  10
MYO1_LYCES
ID   MYO1_LYCES     STANDARD;      PRT;   273 AA.
AC   P54926;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase 1 (EC 3.1.3.25) (IMPase 1) (IMP
DE   1) (Inositol monophosphatase 1).
GN   IMP1.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. VFNT CHERRY;
RX   MEDLINE=96351935; PubMed=8718627;
RA   Gillaspy G.E., Keddie J.S., Oda K., Gruissem W.;
RT   "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RT   by a multigene family.";
RL   Plant Cell 7:2175-2185(1995).
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39444; AAB19030.1; -.
DR   HSSP; P29218; 2HHM.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00378; INOSPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Lithium; Multigene family.
SQ   SEQUENCE   273 AA;  29558 MW;  92D1D532E359E0FD CRC64;


  Query Match             14.3%;  Score 247.5;  DB 1;  Length 273;
  Best Local Similarity   28.9%;  Pred. No. 3.2e-12;
  Matches   80;  Conservative   53;  Mismatches  111;  Indels   33;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           |  : ||  ||  |||::|| | :   :: |     ||  |:  |: : : : : ||:|
Db       8 EEFLGVAVDAAKKAGEIIRKGFHETKHVVHKGQVDLVTETDKACEDLIFNHLKQHFPSHK 67

Qy     136 IFGEENGWRCAENSADF---  -VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           |||     :  : ||       |::||:||| :|: | |     | |     | :|||:
Db      68 FIGEET----SAATGDFDLTDEPTWIVDPVDGTTNFVHGFPSVCVSIGLTIGKIPTVGVV 123

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYL----YTTSPHLFEADAEDAFIRV 245
           ||: | : |::||   |||: | | | : | :: |     ||  :|     |    |:
Db     124 YDPIIDELFTGINGKGAYLNGKPIKVSSQSELVKSLLGTEVGTTRDNL---TVETTTRRI 180

Qy     246 RN---KVKVPLYGCDCYAYAL--LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDW 299
           |   ||:  |  |   |  |  :| | :::    |   |:|      ::: ||| : |
Db     181 NNLLFKVR-SLRMCGSCALDLCWVACGRLELFYLIGYGGPWDVAGGAVIVKEAGGVLFDP 239

Qy     300 RGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           | :  : :|::        |||  :  :  ::||:
Db     240 SGSE--FDITSQR-------VAATNPHLKEAFVEALQ 267


RESULT  11
MYO2_LYCES
ID   MYO2_LYCES     STANDARD;      PRT;   265 AA.
AC   P54927;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase 2 (EC 3.1.3.25) (IMPase 2) (IMP
DE   2) (Inositol monophosphatase 2).
```

```
GN   IMP2.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. VFNT CHERRY;
RX   MEDLINE=96351935; PubMed=8718627;
RA   Gillaspy G.E., Keddie J.S., Oda K., Gruissem W.;
RT   "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RT   by a multigene family.";
RL   Plant Cell 7:2175-2185(1995).
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39443; AAB19029.1; -.
DR   HSSP; P29218; 2HHM.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00378; INOSPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Lithium; Multigene family.
SQ   SEQUENCE   265 AA;  28705 MW;  5369735F30FC884B CRC64;


  Query Match             13.8%;  Score 238;  DB 1;  Length 265;
  Best Local Similarity   30.8%;  Pred. No. 1.7e-11;
  Matches   84;  Conservative   46;  Mismatches  117;  Indels   26;  Gaps    9;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           |  |:||  ||  |||::|  | :   |  |     ||  |:  |  : : : : ||:|
Db       2 EEFVDVAIEAAKKAGEIIRHGFYKSKHIEHKGVVDLVTETDKACEVLIFNHLKQCFPSHK 61

Qy     136 IFGEENGWRCAENSADF------VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
             |||     |  | :|       |::||:||| :|: | |     | |    |||:||:
Db      62 FIGEET---TAAASGNFELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIEKKPVVGVV 118

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT---TSPHLFEADAEDAFI-RV 245
           ||: | :  : |:   |||: | | | : | :| : |    |:      ||    | ||
Db     119 YNPIIDELFTAIYGRGAFLNGKSIRVSSESQLVKALVAT EVGTNRDKAIVDATTGRINRV 178

Qy     246 RNKVK-VPLYGCDCYAYALLASGFVDIV--VESGLKPYI FLSLVPVIEGAGGSITDWRG 302
           ||: : : |        :| | :|:   :| |  |:|  :   ::  ||| : |  |
Db     179 IFKVRSLRMSGSCALNLCGVACGRLDLFYEIEFG-GPWDVAAGALIVIEAGGLVLDPSGS 237

Qy     303 KLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           :  : :||         ||| :| :    ::||
Db     238 E--FDLTARR-------VAATNAHLKDAFINAL 261


RESULT  12
MYOP_MESCR
ID   MYOP_MESCR     STANDARD;      PRT;   270 AA.
AC   O49071;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase (EC 3.1.3.25) (IMPase) (IMP)
DE   (Inositol monophosphatase).
GN   IMP1.
OS   Mesembryanthemum crystallinum (Common ice plant).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
OX   NCBI_TaxID=3544;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nelson D.E., Bohnert H.J.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+) (BY SIMILARITY).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF037220; AAB92418.1; -.
DR   HSSP; P29218; 2HHM.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00378; INOSPHPHTASE.
KW   Hydrolase; Magnesium; Lithium.
SQ   SEQUENCE   270 AA;  29161 MW;  0C045EB33D903A6C CRC64;


  Query Match             13.6%;  Score 235;  DB 1;  Length 270;
  Best Local Similarity   29.0%;  Pred. No. 2.9e-11;
  Matches   72;  Conservative   35;  Mismatches  115;  Indels   26;  Gaps    5;

Qy      85 AADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGEEN--G 142
           ||  ||||:|| |  :  :  |     ||  |:  |: : : : : :| |   |||
Db      17 AAKRAGEVIRKGFYVKKNVEHKGQVDLVTETDKSCEDIIFNCLKQQYPNHKFIGEETTAA 76

Qy     143 WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVD 202
           :  | : :  |::||:||| :|: | |     | |     | :||:  ||: | : ||
Db      77 YGATELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIGKVPTVGVVYNPIMNELFTGVR 136

Qy     203 GKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDA------------FIRVRNKV 249
           :   ||| | | | : |    | |      :   ||              ||:
Db     137 RQGAFLNGVPIHVSSKDELVNCLLVTEVGTKRDKSTVDATTNRINGLLFKVRSIRMAGSC 196

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWRGDKLHWPV 308
           : | |  |         |  |:: |:|    :|  : : ::| ||| | |  |       |
Db     197 ALDLCGIAC--------GRADLMYENGYGGAWDVTAGIVIVEEAGGVIFDPSGKDFDITV 248

Qy     309 T--AESRP 314
           |  | | |
Db     249 TRIAASNP 256


RESULT  13
SUHB_XYLFA
ID   SUHB_XYLFA     STANDARD;      PRT;   275 AA.
AC   Q9PAM0;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
```

```
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR XF2476.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20305717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004055; AAF85274.1; -.
DR   HSSP; P29218; 1IMB.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   275 AA;  30233 MW;  74F622ACC4EE082C CRC64;


  Query Match             13.5%;  Score 234;  DB 1;  Length 275;
  Best Local Similarity   28.4%;  Pred. No. 3.6e-11;
  Matches   73;  Conservative   51;  Mismatches  117;  Indels   16;  Gaps    8;

Qy      79 VEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAI 136
           |  :   :|| :|| || ::    : : :| |       :  |  ||: :|  : :::| : |
Db       6 VNIMVKAARSAGNVLLRHINKLETLHVIQKSRMDYASDVDEMAEKVIVKELKRAYPEYGI 65

Qy     137 FGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
           |||  |    : :   :||:||:|||  ::: | | :   |||: ||:|   ||  |:  |
Db      66 LGEEGG---LQGNHRIMWVIDPLDGTSNYLRGFPHYCISIALVENGEPTDAVIFDPLRNE 122

Qy     197 RWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT-TSP-HLFEADAE----DAFIRVRNKVK 250
           :    |    || ::| | :   |    | |  ||     | |:    || :     ::
Db     123 LFTASRGAGAILNERKIRVANRKDLNGTMLNTGFSPRERSRAHAQLKCVDALLMQAEDIR 182

Qy     251 VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
           |      | :| |  |   |:|:| :|  : : ::  ||| : |::|     |
Db     183 --RSGSAALDLAYVACGRADAYFEAGIKVWDVAAGMLLVREAGGYVCDFKGADA--PRMD 238

Qy     311 ESRPTSFNVVAAGDARV 327
           :  | |  :| ||: :|
Db     239 DKGPESCQLV-AGNIKV 254


RESULT  14
SUHB_THEMA
ID   SUHB_THEMA     STANDARD;      PRT;   256 AA.
AC   O33832;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR TM1415.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogales; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=98386670; PubMed=9720201;
RA   Liebl W., Brem D., Gotschlich A.;
RT   "Analysis of the gene for beta-fructosidase (invertase, inulinase) of
RT   the hyperthermophilic bacterium Thermotoga maritima, and
RT   characterisation of the enzyme expressed in Escherichia coli.";
RL   Appl. Microbiol. Biotechnol. 50:55-64(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99437848; PubMed=10508089;
RA   Chen L., Roberts M.F.;
RT   "Characterization of a tetrameric inositol monophosphatase from the
RT   hyperthermophilic bacterium Thermotoga maritima.";
RL   Appl. Environ. Microbiol. 65:4559-4567(1999).
CC   -!- FUNCTION: Displays a 20-fold higher rate of hydrolysis of the D
CC       isoform of inositol 1-phosphate than of the L isoform.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBUNIT: Homotetramer.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 22.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send a email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AJ001073; CAA04517.1; -.
DR   EMBL; AE001794; AAD36486.1; ALT_FRAME.
DR   HSSP; P29218; 2HHM.
DR   TIGR; TM1415; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00378; INOSPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   256 AA;  28647 MW;  550C15FEEA50B8DC CRC64;


  Query Match             13.3%;  Score 229.5;  DB 1;  Length 256;
  Best Local Similarity   28.7%;  Pred. No. 7.3e-11;
  Matches   68;  Conservative   44;  Mismatches  104;  Indels   21;  Gaps    7;

Qy     100 RVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTAIFGEENGWRCAENSADFVWVLD 157
           ||: ::|:      ||  |||||:  :|  | | ||   |  ||  :        | :|:::|
 b      24 RVDNVEKKTGFKDIVTEIDREAQRMIVDEIRKFFPDENIMAEEGIF----EKGDRLWIID 79

Qy     158 PIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQEISVRS 217
           ||||| :|: | | |   :| : ||:  :|||:  | | |     :|     ||: | | |
Db      80 PIDGTINFVHGLPNFSISLAYVENGEVKLGVVHAPALNETLYAEEGSGAFFNGERIRV-S 138

Qy     218 CNLLAQAYLYTTSPHLFEADAEDAFIRVRNK--VKVPLYGCDCYAYALLAS FVDIVVES 275
           |   :  : :|  ::   |    ||    |   :: : |      | :    ||  |
Db     139 ENASLEECVGSTGSYV---DFTGKFIERMEKRTRRIRILGSAALNAAYVGAGRVDFFVTW 195

Qy     276 GLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEAL 332
           : |:|  : : ::: ||| :||: |          |:   | | : : :  :| | :
Db     196 RINPWDIAAGLIIVKEAGGMVTDFSG--------KEANAFSKNFIFS-NGLIHDEVV 243


RESULT  15
SUHB_RHILO
ID   SUHB_RHILO     STANDARD;      PRT;   266 AA.
AC   Q98F59;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR MLR3921.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP003003; BAB50708.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   266 AA;  28655 MW;  843436E21ED326CF CRC64;


  Query Match             13.2%;  Score 228;  DB 1;  Length 266;
  Best Local Similarity   28.3%;  Pred. No. 1e-10;
  Matches   77;  Conservative   47;  Mismatches  128;  Indels   20;  Gaps    5;

Qy      73 LATERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKS 130
           :|   |: |  :||  ||  | : |   | :::  |     |: |||:||: : : : |:
Db       1 MARSALLNVMVQAAMKAGRSLSRDFGEVQNLQVSLKGPGDYVSQADRKAEDIIFAELSKA 60

Qy     131 FPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVID 190
           | :    || |    |:|    |::||:||| :|: | |||   |||   |: | |||
Db      61 RPGYGFLMEERGAVEGEDS-QHRWIVDPLDGTTNFLHGIPLFAVSIALERQGQIVAGVIY 119

Qy     191 QPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVK 250
           | : | :    |    :| : : |      |    :    |||      :  : :|| :
Db     120 NPAMDELYTTERGGGAFMNDRRLRVAGRIKLVDTVIGCGMPHLGRGHHGNFLVELRNVMA 179

Qy     251 ----VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG--DKL 304
               |    |       | :|:| :|   |:||  :|  : : :|  ||| ::|  |  | |
Db     180 EVSGVRRLGSAALDLAYVAAGRMDGFWETGLSAWDIAAGLLLIREAGGFVSDMDGGQDML 239

Qy     305 HWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                       | ||:  : :  | |::
Db     240 -----------DNGSVVAGNEVIQRALLKAVK 260


Search completed: May 30, 2002, 17:16:51
Job time: 127 sec
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 18, 2002, 22:58:06 ; Search time 52 Seconds
(without alignments)
3496.153 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agctccgagcgtcattctcc.....agccaattttggngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 328367 seqs, 184756068 residues

otal number of hits satisfying chosen parameters: 656734

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56.4	11.5	9953	10	US-09-764-847-1336	Sequence 1336, Ap
	2	56.4	11.5	11189	10	US-09-764-847-1335	Sequence 1335, Ap
	3	55.2	11.2	367	10	US-09-878-574-2578	Sequence 2578, Ap
	4	54.8	11.1	987	10	US-09-804-682-20	Sequence 20, Appl
	5	53.8	10.9	1064	10	US-09-804-682-29	Sequence 29, Appl
	6	52.4	10.7	437	10	US-09-960-352-12835	Sequence 12835, A
	7	51.8	10.5	1614	12	US-10-023-529-45	Sequence 45, Appl
	8	51.8	10.5	1614	12	US-10-023-523-45	Sequence 45, Appl
	9	51.8	10.5	2283	9	US-09-938-842A-1691	Sequence 1691, Ap
	10	51.8	10.5	2283	10	US-09-887-576-645	Sequence 645, App
	11	51.8	10.5	12425	12	US-10-023-529-50	Sequence 50, Appl
	12	51.8	10.5	12425	12	US-10-023-523-50	Sequence 50, Appl
	13	51.2	10.4	5025	10	US-09-960-253-176	Sequence 176, App
	14	49.8	10.1	277	10	US-09-878-574-14649	Sequence 14649, A
	15	49.8	10.1	2561	12	US-10-023-529-48	Sequence 48, Appl
	16	49.8	10.1	2561	12	US-10-023-523-48	Sequence 48, Appl
	17	48.4	9.8	2481	10	US-09-894-998-35	Sequence 35, Appl
c	18	48.2	9.8	326	10	US-09-864-761-17105	Sequence 17105, A
	19	48.2	9.8	474	10	US-09-864-761-11284	Sequence 11284, A
c	20	48.2	9.8	477	10	US-09-864-761-286	Sequence 286, App
c	21	48.2	9.8	477	10	US-09-864-761-287	Sequence 287, App
	22	47.6	9.7	463	10	US-09-864-761-5509	Sequence 5509, Ap
	23	47.6	9.7	152331	9	US-10-095-407-16	Sequence 16, Appl
	24	47.4	9.6	1065	10	US-09-804-682-33	Sequence 33, Appl
c	25	47.2	9.6	936	10	US-09-974-300-836	Sequence 836, App
	26	47.2	9.6	42999	9	US-09-799-462A-17	Sequence 17, Appl
	27	47.2	9.6	42999	9	US-10-125-767-17	Sequence 17, Appl
c	28	46.8	9.5	410	10	US-09-983-965-4888	Sequence 4888, Ap
	29	46.6	9.5	267	10	US-09-864-761-27984	Sequence 27984, A
c	30	46.6	9.5	3809	12	US-10-001-870-68	Sequence 68, Appl
	31	46.4	9.4	4450	10	US-09-764-887-497	Sequence 497, App
c	32	45.8	9.3	404	10	US-09-983-965-5091	Sequence 5091, Ap
c	33	45.8	9.3	57130	10	US-09-835-081-3	Sequence 3, Appli
	34	45.2	9.2	659158	9	US-09-771-208-20	Sequence 20, Appl
c	35	44.8	9.1	501	10	US-09-878-574-4421	Sequence 4421, Ap
	36	44.8	9.1	531	10	US-09-864-761-22281	Sequence 22281, A
	37	44.8	9.1	1602	10	US-09-833-790-417	Sequence 417, App
	38	44.8	9.1	4454	10	US-09-764-887-496	Sequence 496, App
c	39	44.8	9.1	174424	10	US-09-967-768A-314	Sequence 314, App
c	40	44.6	9.1	1203	9	US-09-879-312-1	Sequence 1, Appli
c	41	44.4	9.0	367	10	US-09-878-574-2578	Sequence 2578, Ap
c	42	44.4	9.0	3901	10	US-09-834-975-1043	Sequence 1043, Ap
c	43	44.2	9.0	106	10	US-09-960-352-5165	Sequence 5165, Ap
c	44	44.2	9.0	152331	9	US-10-095-407-16	Sequence 16, Appl
	45	43.8	8.9	266	10	US-09-960-352-3882	Sequence 3882, Ap

ALIGNMENTS

RESULT 1
US-09-764-847-1336
; Sequence 1336, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1336
; LENGTH: 9953
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1336

Query Match 11.5%; Score 56.4; DB 10; Length 9953;
Best Local Similarity 58.2%; Pred. No. 0.0002;
Matches 99; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy      21 GCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCT 80
           ||   ||  ||  |  ||| || ||  || ||||||||  || |||| |      |||||
Db    3991 GCTTTGAGCTTGGAGGCTTCCTGCCCGCAGCCGCCGCCTCCTTCTCCTCCTCCTCCTCCT 4050

Qy      81 CTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTA 140
           | |||||||||  ||  |||  || ||||  ||| | ||  ||||  | |  |    |||
Db    4051 CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCCTCCTCCCTCCTCCTCCTC 4110

Qy     141 ACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCC 190
             | ||  ||| | ||||||||  |  |   |  | | ||    ||  ||
Db    4111 CCTCCTCCTCCCTCCTCCTCCCTCCTCCTCCTCCTCCTCCTCCTTCCCCC 4160
```

RESULT 2
US-09-764-847-1335
; Sequence 1335, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

```
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1335
;   LENGTH: 11189
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1335

  Query Match             11.5%;  Score 56.4;  DB 10;  Length 11189;
  Best Local Similarity   58.2%;  Pred. No. 0.0002;
  Matches   99;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy      21 GCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCT 80
           ||   ||  ||  |  ||| || ||  || ||||||||  || |||| |      |||||
Db    5227 GCTTTGAGCTTGGAGGCTTCCTGCCCGCAGCCGCCGCCTCCTTCTCCTCCTCCTCCTCCT 5286

Qy      81 CTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTA 140
           | |||||||||  ||  |||  || ||||  ||| | ||  ||||  | |  |    |||
Db    5287 CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCCTCCTCCCTCCTCCTCCTC 5346

Qy     141 ACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCC 190
           | ||  |||  | ||||||||  |  |   |  |  | ||    ||  ||
Db    5347 CCTCCTCCTCCCTCCTCCTCCCTCCTCCTCCTCCTCCTCCTTCCCCC 5396


RESULT 3
US-09-878-574-2578
; Sequence 2578, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2578
;   LENGTH: 367
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: LIB3028-023-Q1-B1-C4
US-09-878-574-2578

  Query Match             11.2%;  Score 55.2;  DB 10;  Length 367;
  Best Local Similarity   51.3%;  Pred. No. 0.00015;
  Matches  176;  Conservative    0;  Mismatches  164;  Indel ·    3;  Gaps    2;

Qy      59 GACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCC 118
           | |||  || ||   | | || ||| |||   |  |||||| |||||| |  ||| | ||
Db      11 GCCTCAGCCGCATCCCCCACCACTTTCTCTCTCCTCC.:TCCGCCACCCCTTCCTCCTCC 70

Qy     119 GGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCG-CCTC 177
           |  |||| || |   | |   | ||| ||| ||| | |  | |  |  | ||  || |
Db      71 TCCACCGCCACCTCCGCCTCCCCCTCCCCCTCGCGCGTGCAGCTCACCTCCTCCCTCCGC 130

Qy     178 GCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCG--CCAGCCGATGAGTACGGTTAGG 235
           |||||     | ||| | || |    ||    ||||  || || | |     ||  |
Db     131 GCCCGCCCGCTGCGCCGCCTGGGCTTCGCCGCGGCGGACCCGCTGCTCGCGGCGAGTGTA 190

Qy     236 GCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACG 295
           ||  ||  ||  || | || |   ||  |  ||   |  ||   ||||| | || | | |
Db     191 GCGGCGCGCGTGGCGGCGGTGAGAGGAGGAGGAAAGGGCGCGAGGGGGGTGGTGTCCATG 250

Qy     296 GAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAG 355
           | |  |  |   | || ||  |   | |||||||||      |||| ||  | ||
Db     251 GCGAAGAAGAGCGTGGGGGACCTGAGCGCGGCGGATCTGAATGGGAAGAAGGTCTTCGTC 310

Qy     356 TACTTCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACA 398
               |  ||     ||   | ||  ||||||  || ||  ||
Db     311 AGGGCCGACCTCAAAGTCCCGCTCGACGACAACCAGAACATCA 353


RESULT 4
US-09-804-682-20
; Sequence 20, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
;  APPLICANT: Kinders, Robert
;  APPLICANT:  Corey, Michael J.
;  TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
;  TITLE OF INVENTION:  ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
;  TITLE OF INVENTION:  MODULATING THE SAME
;  FILE REFERENCE: 130001.406
;  CURRENT APPLICATION NUMBER: US/09/804,682
;  CURRENT FILING DATE:  2001-03-12
;  NUMBER OF SEQ ID NOS: 174
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
;   LENGTH: 987
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 5, 7, 12, 13, 14, 15, 26, 39, 52, 87, 120, 121, 122, 123,
;   LOCATION: 124, 125, 128, 130, 132, 133, 137, 141, 151, 152, 154, 160,
;   LOCATION: 167, 174, 182, 184, 187, 189, 190, 192, 198, 200, 205, 219,
;   LOCATION: 220, 221, 222, 225, 231, 233, 237, 238, 242, 246
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: 250, 251, 2  , 253, 259, 264, 266, 2 9, 271, 277, 278, 281,
;   LOCATION: 298, 300, 3  , 304, 305, 308, 309, 3 5, 322, 324, 329, 334,
;   LOCATION: 346, 347, 349, 351, 356, 358, 359, 364, 365, 375, 376, 377,
;   LOCATION: 384, 392, 395, 405, 428, 432, 442, 446, 449, 454, 461
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: 462, 464, 473, 479, 484, 485, 488, 491, 503, 512, 518, 539,
;   LOCATION: 540, 543, 544, 555, 556, 561, 565, 569, 573, 575, 577, 578,
;   LOCATION: 579, 580, 582, 584, 589, 591, 600, 602, 614, 616, 621, 636,
;   LOCATION: 649, 650, 663, 665, 670, 678, 684, 694, 705, 709, 711
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: 712, 717, 718, 720, 723, 734, 737, 755, 756, 760, 762, 764,
;   LOCATION: 765, 773, 785, 789, 797, 798, 799, 800, 804, 807, 822, 823,
;   LOCATION: 825, 826, 830, 831, 837, 848, 852, 868, 875, 887, 893, 904,
;   LOCATION: 907, 908, 911, 932, 939, 946, 954, 962, 971, 974, 976
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: 978, 980, 981
;   OTHER INFORMATION: n = A,T,C or G
US-09-804-682-20

  Query Match             11.1%;  Score 54.8;  DB 10;  Length 987;
  Best Local Similarity   52.9%;  Pred. No. 0.00025;
  Matches   83;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy      41 CTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCG 100
           | | |  |  || |  ||  | || ||||    | | || |  || | |||  || | |
Db     618 CCCNCCCCCCCCCCCTCCCNTCCCCCCCCCCCCNNCCCCCCCCCCCCCNCNTCCCNCCCCACC 677

Qy     101 GCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTC 160
           || || | ||||  | |  | || || ||     ||||  | ||  | ||| ||| ||
Db     678 NCCCCCNCCCCCTCTCNCCTCATCCCCNCCCNCNNCCCTNNCNCCNCCCCTCCCCTNCTN 737

Qy     161 CGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCT 197
           |  |||| | ||  | |  ||    |  || |  | |
Db     738 CTACTCCCCCCCCTCCCNNCCCNCNCNNCCTCCCTNT 774
```

```
RESULT 5
US-09-804-682-29
; Sequence 29, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
;  APPLICANT: Kinders, Robert
;  APPLICANT:  Corey, Michael J.
;  TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
;  TITLE OF INVENTION:  ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
;  TITLE OF INVENTION:  MODULATING THE SAME
;  FILE REFERENCE: 130001.406
;  CURRENT APPLICATION NUMBER: US/09/804,682
;  CURRENT FILING DATE:  2001-03-12
;  NUMBER OF SEQ ID NOS: 174
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 1064
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 5, 6, 16, 21, 24, 25, 33, 39, 72  110, 209, 214, 231, 232,
;   LOCATION: 235, 237, 238, 244, 245, 246, 25 , 282, 292, 297, 306, 319,
;   LOCATION: 321, 323, 330, 334, 340, 349, 354, 355, 363, 372, 376, 378,
;   LOCATION: 397, 405, 432, 437, 454, 455, 457, 458, 459, 468, 470
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: 485, 487, 488, 494, 496, 499, 511, 524, 527, 552, 557, 562,
;   LOCATION: 583, 600, 611, 613, 623, 624, 652, 654, 674, 681, 687, 691,
;   LOCATION: 694, 701, 713, 716, 720, 721, 725, 731, 734, 735, 739, 743,
;   LOCATION: 744, 781, 782, 785, 789, 799, 803, 821, 823, 847, 852
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: 858, 878, 884, 886, 896, 897, 901, 917, 926, 932, 939, 948,
;   LOCATION: 957, 961, 965, 981, 991, 993, 1001, 1002, 1005, 1011, 1018,
;   LOCATION: 1043, 1047, 1049, 1051, 1054, 1056
;   OTHER INFORMATION: n = A,T,C or G
US-09-804-682-29

  Query Match            10.9%;  Score 53.8;  DB 10;  Length 1064;
  Best Local Similarity   50.8%;  Pred. No. 0.00044;
  Matches   91;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
           ||||   |     | ||  |||| |      ||||  | |||  |   | |   |
Db     677 CTCCNCCCCCNCCTNTCNCCCCCCNCCCCCTCACCTNC '' 'CCNNCTCNCCCTCNCCNNC 736

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
           || |||   | | | || || || || ||  || |||  || ||   | |    ||  ||
Db     737 CCNCCCNNCT CCCCCCCCCTACCCCCCCCTTCCCCTCCCCCCCCNNCCNCCCNTCCCCCC 796

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCC 181
           | ||  | |   ||||  ||||| | | | | ||  || || ||   || ||| |  ||
Db     797 CCNCCCNCCCCCCCCCTCCCCCCCNCNCCCCCTTCTCCCTCCCCCCTTCCNCCCCNTCC 855


RESULT 6
US-09-960-352-12835
; Sequence 12835, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12835
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 55-LIB2809-007-Q1-E1-F4
US-09-960-352-12835

  Query Match            10.7%;  Score 52.4;  DB 10;  Length 437;
  Best Local Similarity   55.5%;  Pred. No. 0.00073;
  Matches  101;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy      16 TCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCT 75
           ||||| ||||  |   |   |||  | |||  |  || || ||  ||||||||
Db     227 TCTCCT CCCCCTCTCCTCCCCCTCCCCCCCCCCCCCCCCCCCCCCCTCCTCCCTCCCCTC 286

Qy      76 CTCCTCTTCCTCCTC ACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAA 135
           | |  |  || || || ||   ||  |  || | ||| ||| || || | ||  | |
Db     287 CCCTCCCCCCCCCCCCTCCTTTTCTCCTTCCCCACCCCTCCACGCCCCCCCCGTCCCCTT 346

Qy     136 CCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGT 195
           |||   ||||| ||| | ||  | || |||||   || ||||  ||       || | |
Db     347 CCCCCCCCCCCTCTCCCTCCCTCCCCCCCTCCCTCCCCCCTCCTCCCCCCTCCCCCCTCT 406

Qy     196 CT 197
           ||
Db     407 CT 408


RESULT 7
US-10-023-529-45
; Sequence 45, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,529
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 1614
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1614)
US-10-023-529-45

  Query Match            10.5%;  Score 51.8;  DB 12;  Length 1614;
  Best Local Similarity   49.2%;  Pred. No. 0.0014;
  Matches  189;  Conservative    0;  Mismatches  193;  Indels    2;  Gaps    2;

Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
           | |||  ||       ||  | ||| |       ||  |   ||     | |||||| |
Db     404 CACCGCCCGCGCCCGTCGCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCG 463

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
```

```
Db    464 CCGCCACAGCGCCCCCCTCGCCTGGCCCCGCGCAGCCGGGC-CCCCGCGCGCGCAGCGGGCC 522

Qy    123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182

Db    523 GCGCCCCTGGCCGCGCCGCCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCC 582

Qy    183 TGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGT 242

Db    583 GGCCCGCGCCGCGCCCCCCCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCG 642

Qy    243 TCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGAC-GGAGCGG 301

Db    643 CCACAGCCGCCGGCGCCGCCACAGCAGCAGCAGCCGCCGCCGCCGCCGCAGCCACAGCCGCCG 702

Qy    302 CTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTC 361

Db    703 CCGGAGGGGGGCGCGGTGCGGGCCGGCGGCGCGGCGCGGCCCGTGAGCCTGCGGGAAGTC 762

Qy    362 CGCCAGCGGGTTGAGATCATCGAC 385

Db    763 GTGCGCTACCTCGGGGGCAGCGGC 786


RESULT 8
US-10-023-523-45
; Sequence 45, Application US/10023523
; Patent No. US20020152485A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,523
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: US/09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;  LENGTH: 1614
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(1614)
US-10-023-523-45

  Query Match           10.5%;  Score 51.8;  DB 12;  Length 1614;
  Best Local Similarity  49.2%;  Pred. No. 0.0014;
  Matches  189;  Conservative    0;  Mismatches  193;  Indels    2;  Gaps    2;

Qy      3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62

Db    404 CACCGCCCGCGCCCGTCGCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCG 463

Qy     63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122

Db    464 CCGCCACAGCGCCCCCCTCGCCTGGCCCCGCGCAGCCGGGC-CCCCGCGCGCGCAGCGGGCC 522

Qy    123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182

Db    523 GCGCCCCTGGCCGCGCCGCCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCC 582

Qy    183 TGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGT 242

Db    583 GGCCCGCGCCGCGCCCCCCCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCG 642

Qy    243 TCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGAC-GGAGCGG 301

Db    643 CCACAGCCGCCGGCGCCGCCACAGCAGCAGCAGCCGCCGCCGCCGCCGCAGCCACAGCCGCCG 702

Qy    302 CTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTC 361

Db    703 CCGGAGGGGGGCGCGGTGCGGGCCGGCGGCGCGGCGCGGCCCGTGAGCCTGCGGGAAGTC 762

Qy    362 CGCCAGCGGGTTGAGATCATCGAC 385

Db    763 GTGCGCTACCTCGGGGGCAGCGGC 786


RESULT 9
US-09-938-842A-1691
; Sequence 1691, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINI
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1691
;  LENGTH: 2283
;  TYPE: DNA
;  ORGANISM: Arabidopsis thaliana
US-09-938-842A-1691

  Query Match           10.5%;  Score 51.8;  DB 9;  Length 2283;
  Best Local Similarity  51.5%;  Pred. No. 0.0016;
  Matches  119;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy     22 CCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTC 81

Db   1295 CTCCGCCTGTTTATTCTCCCCCTCCTCCACCGCCACCACCTCCTCCGGTATATTCTCCTC 1354

Qy     82 TTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAA 141

Db   1355 CACCACCACCGCCCCCACCGCCTCCTCCGCCAGTATATTCTCCTCCACCACCACCACCGC 1414

Qy    142 CCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAG 201

Db   1415 CCCCACCGCCTCCTCCGCCAGTATATTCTCCGCCACCACCATCGCCGCCTCCACCGCCTC 1474

Qy    202 CGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGG 252

Db   1475 CGCCAGTCTACTCTCCCCCACCACCACCACCGCCTCCACCTCCTCCGCCGG 1525


RESULT 10
US-09-887-576-645
; Sequence 645, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Budworth, P.
;  APPLICANT:  Brown, D.
;  APPLICANT:  Chang, H.
;  APPLICANT:  Zhu, T.
;  APPLICANT:  Han, B.
;  APPLICANT:  Wang, X.
;  APPLICANT:  Cooper, Bret
;  TITLE OF INVENTION: Promoters for regulation of plant expression
;  FILE REFERENCE: 1360.001US1
;  CURRENT APPLICATION NUMBER: US/09/887,576
;  CURRENT FILING DATE:  2001-06-25
;  PRIOR APPLICATION NUMBER: US 60/213,848
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/214,087
;  PRIOR FILING DATE: 2000-06-23
;  PRIOR APPLICATION NUMBER: US 60/258,692
;  PRIOR FILING DATE: 2000-12-29
;  NUMBER OF SEQ ID NOS: 875
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 645
;   LENGTH: 2283
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-887-576-645

  Query Match             10.5%;  Score 51.8;  DB 10;  Length 2283;
  Best Local Similarity   51.5%;  Pred. No. 0.0016;
  Matches  119;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy      22 CCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTC 81
           | ||| |  || |  ||  | | |  | |||||| ||  |||||||   ||  |||||||
Db    1295 CTCCGCCTGTTTATTCTCCCCCTCCTCCACCGCCACCACCTCCTCCGGTATATTCTCCTC 1354

Qy      82 TTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAA 141
             || || || ||| | | ||| ||||  ||        |  | || || | | | |
Db    1355 CACCACCACCGCCCCCACCGCCTCCTCCGCCAGTATATTCTCCTCCACCACCACCACCGC 1414

Qy     142 CCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAG 201
           ||||  | |   |  |  | | |   | |||||  |  || |  |  || |   |
Db    1415 CCCCACCGCCTCCTCCGCCAGTATATTCTCCGCCACCACCATCGCCGCCTCCACCGCCTC 1474

Qy     202 CGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGG 252
           ||| |||   | |   || |  |  ||        || | | | |||| ||
Db    1475 CGCCAGTCTACTCTCCCCCACCACCACCACCGCCTCCACCTCCTCCGCCGG 1525


RESULT 11
US-10-023-529-50
; Sequence 50, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,529
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
;   LENGTH: 12425
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-023-529-50

  Query Match             10.5%;  Score 51.8;  DB 12;  Length 12425;
  Best Local Similarity   49.2%;  Pred. No. 0.0025;
  Matches  189;  Conservative    0;  Mismatches  193;  Indels    2;  Gaps    2;

Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
           | |||  ||       ||  | ||| |       ||  |   ||     | |||||| |
Db    3235 CACCGCCCGCGCCCGTCGCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCG 3294

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
           || || ||  || | ||||  |   | || | ||  ||| | || ||| |    | ||||
Db    3295 CCGCCACAGCGCCCCCCTCGCCTGGCCCCGCGCAGCCGGGC-CCCCGCGCGCAGCGGGCC 3353

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182
           |||| | |    | |   | |||||  |  ||| |  || |  | | | |||| |   |
Db    3354 GCGCCCCTGGCCGCGCCGCCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCC 3413

Qy     183 TGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGT 242
           | |  |  ||  |      |||  |  | ||||  |||       | | |   ||| |
Db    3414 GGCCCGCGCCGCGCCCCCCCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCG 3473

Qy     243 TCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGAC-GGAGCGG 301
           | | || |  |  ||||  |      ||| || || |  |  |     || || |  || |
Db    3474 CCACAGCCGCCGGCGCCGCCACAGCAGCAGCAGCCGCCGCCGCCGCAGCCACAGCCGCCG 3533

Qy     302 CTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTC 361
           |  | || ||  |||   ||||| |  | ||| ||  ||   |||  |  |  | |  ||
Db    3534 CC GAGGGGGGCGCGGTGCGGGCCGGCGGCGCGGCGCGGCCCGTGAGCCTGCGGGAAGTC 3593

Qy     362 CGCCAGCGGGTTGAGATCATCGAC 385
               |      | | |  || || |
Db    3594 GTGCGCTACCTCGGGGGCAGCGGC 3617


RESULT 12
US-10-023-523-50
; Sequence 50, Application US/10023523
; Patent No. US20020152485A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,523
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: US/09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
;   LENGTH: 12425
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-023-523-50

  Query Match             10.5%;  Score 51.8;  DB 12;  Length 12425;
```

```
  Best Local Similarity   49.2%;  Pred. No. 0.0025;
  Matches  189;  Conservative    0;  Mismatches  193;  Indels    2;  Gaps    2;

Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
           | |||  ||       ||  | ||| |        ||  |   ||      | |||||| |
Db    3235 CACCGCCCGCGCCCGTCGCCGCCGCCGCCCCGGCCCGGGCGCCCCGCGCGGCCGCCGCCG 3294

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
           || || ||  || | ||||  |   | || | ||  ||| | || ||| |    | ||||
Db    3295 CCGCCACAGCGCCCCCCTCGCCTGGCCCCGCGCAGCCGGGC-CCCCGCGCGCAGCGGGCC 3353

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182
            |||| | |    | |   | ||||| |  ||| |  || |  | | | |||| |   |
Db    3354 GCGCCCCTGGCCGCGCCGCCGCCCGCGCCAGCCGCTCCCCCGGCGGTGGCGCCCCCGGCC 3413

Qy     183 TGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGT 242
           | |  |  ||  |     |||  |  | ||||  |||       | | |   ||| |
Db    3414 GGCCCGCGCCGCGCCCCCCCGCCCGCCGTCGCCGCCCGGGAGCCGCCGCTGCCGCCGCCG 3473

Qy     243 TCGCCCTGGGGCGGCCGGCCGGAGAGCTGCGGCA TGGGGGAG TGGCGAC-GGAGCGG 301
           | | || |  |  |||| |     ||| || || |  |  |     || || |  || |
Db    3474 CCACAGCCGCCGGCGCCGCCACAGCAGCAGCAGCCGCCGCCGCCGCAGCCACAGCCGCCG 3533

Qy     302 CTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTC 361
           | || || ||  |||   ||||| |  | ||| ||  ||   |||  |  |  | |  ||
Db    3534 CCGGAGGGGGGCGCGGTGCGGGCCGGCGGCGCGGCGCGGCCCGTGAGCCTGCGGGAAGTC 3593

Qy     362 CGCCAGCGGGTTGAGATCATCGAC 385
               |      | | |  || || |
Db    3594 GTGCGCTACCTCGGGGGCAGCGGC 3617


RESULT 13
US-09-960-253-176
; Sequence 176, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
;  APPLICANT: Benson, Darin R.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Lodes, Michael J.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.556
;  CURRENT APPLICATION NUMBER: US/09/960,253
;  CURRENT FILING DATE:  2001-09-20
;  NUMBER OF SEQ ID NOS: 187
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 176
;   LENGTH: 5025
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-960-253-176

  Query Match             10.4%;  Score 51.2;  DB 10;  Length 5025;
  Best Local Similarity   53.8%;  Pred. No. 0.0027;
  Matches  128;  Conservative    0;  Mismatches  108;  Indels    2;  Gaps    1;

Qy     111 CCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCG 170
           | |||||   || ||   |  |  ||||   | |||| ||| | ||| ||||| || |
Db     502 CTTTCCCGCCCCCCGTTCCTCCCGACCCCTCCACCCGCCCTCCCTTCCCCCGCCGCCCCT 561

Qy     171 CCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGG 230
           || ||||  ||  |  |   |||| ||   ||| | | ||| |   || |  | |  |||
Db     562 CCTCCTCCTCCCCGGAGGGGGCGGGCTCCGGCGGGTGCGGGGGTGGGCGGGCGGGGCCGG 621

Qy     231 TTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAG--TGGGGGAGTT 288
              ||  |||   |  |   |  |  | | |||| || | || || |   || | | ||
Db     622 GGGTGGGGTCGGCGGGGGACCGTCCCCCGGACCGGCGACCGGCCGCCGCCGGGCGCATTT 681

Qy     289 GGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTG 346
             | ||| ||| | |    | |   | | |   ||  | | | || ||| |||||
Db     682 CCAGGCGGTGCGCCGCGACCGGCTCCGGGACGGCTGGCAAGGCCCGGCGGGGAAGGTC 739
```

```
RESULT 14
US-09-878-574-14649
; Sequence 14649, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 14649
;   LENGTH: 277
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: 701069068H1
US-09-878-574-14649

  Query Match             10.1%;  Score 49.8;  DB 10;  Length 277;
  Best Local Similarity   59.1%;  Pred. No. 0.0026;
  Matches  101;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy     297 AGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGT 356
           | ||  | |  |||||  |   |    |  || || |||||  || || ||  ||   || |
Db     106 ATCGCTTCGCCGAGGTCGGTAACAAAGTCGCCGATGCTGCCGGAGAAGTTATCCGCAAAT 165

Qy     357 ACTTCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAG 416
           |||||  |  |     | ||  | ||  ||||| | || | ||||||  || || |||||||
Db     166 ACTTCAGAAAAAACTTCGACGTTATTCACAAACATGATCTCAGTCCAGTAACCATTGCAG 225

Qy     417 ATAGAAGAAGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCT 467
           |  ||   || || || || ||||  ||| | |||||    |  | || |
Db     226 CT-CAATCTGCTGAGGAGGCTATGGTTTCAATCATACTAGCAATTTCCCTT 275


RESULT 15
US-10-023-529-48
; Sequence 48, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
;  APPLICANT: Lees, Ann M.
;  APPLICANT:  Lees, Robert S.
;  APPLICANT:  Law, Simon W.
;  APPLICANT:  Arjona, Anibal A.
;  TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;  TITLE OF INVENTION:  PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;  TITLE OF INVENTION:  ATHEROSCLEROSIS
;  FILE REFERENCE: 10797-004001
;  CURRENT APPLICATION NUMBER: US/10/023,529
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: 09/616,289
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 09/517,849
;  PRIOR FILING DATE: 2000-03-02
;  PRIOR APPLICATION NUMBER: US 08/979,608
;  PRIOR FILING DATE: 1997-11-26
;  PRIOR APPLICATION NUMBER: US 60/031,930
;  PRIOR FILING DATE: 1996-11-27
;  PRIOR APPLICATION NUMBER: US 60/048,547
;  PRIOR FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 53
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48
;   LENGTH: 2561
;   TYPE: DNA
;   ORGANISM: Oryctolagus cuniculus
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (246)...(1895)
US-10-023-529-48

  Query Match             10.1%;  Score 49.8;  DB 12;  Length 2561;
  Best Local Similarity   52.1%;  Pred. No. 0.0048;
  Matches  158;  Conservative    0;  Mismatches  142;  Indels    3;  Gaps    2;

Qy      41 CTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCG 100
           | | || |  |||||||||   || || ||  || | ||||  ||  | || | ||  ||
Db     693 CGCGCGGCCGCCGCCGCCGCTGCCGCCACAGCGCCCCCCTCGCCCGGCCCCGCGCAGCCG 752

Qy     101 GCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTC 160
           | | || ||| |    | |||| |||| | |    | |   | ||||| | |||| |  |
Db     753 GGC-CCCCGCGCGCAGCGGGCCGCGCCCCTGGCCGCGCCGCCGCCCGCGCCCGCCGCTCC 811

Qy     161 CGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCG 220
           | |  | ||| |||| |   |  | |  |  ||  |     |||  |  | ||||  |
Db     812 CCCGGCGGCGGCGCCCCCGGCCGGCCCGCGCCGCGCCCCCCCGCCCGCCGCCGCCGTCGC 871

Qy     221 ATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTG 280
                 ||    ||  |   ||||||    || ||||||        | || |||  |
Db     872 CGCCCGGGAGTCGCCGCTGCCGCCGCCGCCACAGCCGCCGGCGCCGCCACAGCAGCA--G 929

Qy     281 GGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGG 340
             | ||  | || ||   | || |  | ||  | | || | | || |||| |  |  | |
Db     930 CAGCAGCCGCCGCCGCCACCGCCGCCGCAGCAGCCACAGCCGCCGCCGGAGGGGGGCGCG 989

Qy     341 GAG 343
           | |
Db     990 GCG 992


Search completed: November 19, 2002, 00:27:47
Job time : 87 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 18, 2002, 19:29:19 ; Search time 2102 Seconds
(without alignments)
3790.761 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agctccgagcgtcattctcc.....agccaattttggngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

otal number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	297	60.4	505	12	BG103856	BG103856 RHIZ2_37_
2	297	60.4	533	12	BG053327	BG053327 RHIZ2_26_
3	177.2	36.0	444	14	BQ240879	BQ240879 TaE05011G
4	176	35.8	588	13	BI778441	BI778441 EBro07_SQ
5	175	35.6	863	13	BI951067	BI951067 HVSME1002
6	171.2	34.8	598	13	BM370373	BM370373 EBro08_SQ
7	168.4	34.2	483	9	AU093884	AU093884 AU093884
8	168.4	34.2	864	10	BE418829	BE418829 SCL081.H0
9	161	32.7	766	12	BF631396	BF631396 HVSMEb001
10	155.2	31.5	443	14	BQ467759	BQ467759 HS04M21r
11	148.8	30.2	404	10	BE498543	BE498543 WHE0971_F
12	145	29.5	667	13	BI959675	BI959675 HVSMEn002
13	141.8	28.8	631	12	BG414965	BG414965 HVSMEk000
14	119.2	24.2	759	14	BQ801093	BQ801093 WHE2810_C
15	115.2	23.4	745	14	BQ608457	BQ608457 BRY_4362
16	90	18.3	528	10	BE443955	BE443955 WHE1123_D
c 17	83.6	17.0	974	17	BH157536	BH157536 ENTRL88TR
18	79.4	16.1	974	17	BH157536	BH157536 ENTRL88TR
19	78.8	16.0	341	13	BI120461	BI120461 F015P69Y
20	78.4	15.9	511	10	AV527020	AV527020 AV527020
21	78.4	15.9	528	10	AV551269	AV551269 AV551269
22	78.4	15.9	622	9	AU239852	AU239852 AU239852
23	78	15.9	322	12	BF428639	BF428639 WHE1411_E
24	76.4	15.5	783	10	BE641140	BE641140 Cri2_2_G0
c 25	74.8	15.2	840	17	AG043467	AG043467 Pan trogl
26	74.6	15.2	724	17	AZ197686	AZ197686 SP_1035_B
c 27	74.4	15.1	939	17	AG043613	AG043613 Pan trogl
c 28	72	14.6	724	17	AZ197686	AZ197686 SP_1035_B
c 29	71.4	14.5	843	17	CNS00CS1	AL059666 Drosophil
30	71	14.4	1212	14	BQ069498	BQ069498 AGENCOURT
31	70.8	14.4	1192	13	BM556419	BM556419 AGENCOURT
c 32	70.2	14.3	1160	17	AG043473	AG043473 Pan trogl
33	70	14.2	624	9	AI727969	AI727969 BNLGHi984
c 34	69.8	14.2	1168	14	BM809732	BM809732 AGENCOURT
35	69	14.0	421	17	BH254057	BH254057 SALK_0159
36	69	14.0	423	14	BQ094569	BQ094569 san48h09.
37	69	14.0	505	12	BG651294	BG651294 sad99g06.
38	69	14.0	551	13	BI972349	BI972349 sag90c12.
39	69	14.0	553	14	BM731926	BM731926 sal86e03.
40	69	14.0	584	13	BM524744	BM524744 sal19c08.
41	69	14.0	609	12	BG238817	BG238817 sab53e03.
42	69	14.0	627	10	AW760124	AW760124 s158d05.y
43	68.4	13.9	1101	17	CNS00397	AL063912 Drosophil
c 44	67.6	13.7	729	17	CNS03JVQ	AL247391 Tetraodon
45	67.4	13.7	945	17	CNS0076K	AL066880 Drosophil

ALIGNMENTS

RESULT 1
BG103856
LOCUS BG103856 505 bp mRNA linear EST 30-JAN-2001
DEFINITION RHIZ2_37_F05.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA sequence.
ACCESSION BG103856
VERSION BG103856.1 GI:12618689
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 505)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 495
POLYA=No.
FEATURES Location/Qualifiers

```
     source          1. .505
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      104 a    131 c    170 g    100 t
ORIGIN

  Query Match             60.4%;  Score 297;  DB 12;  Length 505;
  Best Local Similarity   89.4%;  Pred. No. 4.5e-50;
  Matches  329;  Conservative    0;  Mismatches   38;  Indels    1;  Gaps    1;

Qy     125 GCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTG 184
           |||||| |||||||||||||||||| |||||| |||||||||| ||||| |||| ||||
Db       9 GCCTCCACAAACCCTAACCCCCGCTCCCGCCTCGTCCGCCTCCGAGCCGCATCGCGCGTG 68

Qy     185 TCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTC 244
           ||||||||| ||| | |||| | ||||||| ||||||| || |||| ||||| |||| |
Db      69 CCGTCCGCGGCCTTAAACGCGCGCGGGCGCCCGCCGATGTGTTCGGTCAGGGCTTCGTCC 128

Qy     245 GCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTG 304
           |||||||| |||| |||| || |||| ||||||||||||||||||||||||||| ||||||
Db     129 GCCGCTGGTGCGGGCGGCTGGGGAGCGGCGGCAGTGGGGGAGTTGGCGACGGAACGGCTG 188

Qy     305 GTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGC 364
           |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||||
Db     189 GTGGAGGTGGCGCAGCGGGCGGCAGACGCGGCCGGGGAGGTGCTCAGGAAGTACTTCCGC 248

Qy     365 CAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGA 424
           |||||||||||||||||||||| |||||| |||| |||||||| |||||||||| |||
Db     249 CAGCGGGTTGAGATCATCGACAAGGAGGACAACAGCCCTGTTACGATTGCAGATAG-AGA 307

Qy     425 AGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGG 484
           |||||||||||||||| |||||||||||||||||||||||||||| |||| ||||||
Db     308 AGCAGAAGAAGCAATGGTGTCAGTTATACTGAAGAGCTTCCCTACTCATGCCATTTTTGG 367

Qy     485 NGAAGAGA 492
           || ||||
Db     368 TGAGGAGA 375


RESULT 2
BG053327
LOCUS       BG053327                 533 bp    mRNA    linear   EST 25-JAN-2001
DEFINITION  RHIZ2_26_A12.b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
  CESSION   BG053327
.ERSION     BG053327.1  GI:12508897
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Seq primer: JEN REV
            High quality sequence stop: 507
            POLYA=No.
FEATURES             Location/Qualifiers
```

```
     source          1. .533
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      111 a    137 c    176 g    109 t
ORIGIN

  Query Match             60.4%;  Score 297;  DB 12;  Length 533;
  Best Local Similarity   89.4%;  Pred. No. 4.6e-50;
  Matches  329;  Conservative    0;  Mismatches   38;  Indels    1;  Gaps    1;

Qy     125 GCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTG 184
           |||||| |||||||||||||||||| |||||| |||||||||| ||||| |||| ||||
Db       9 GCCTCCACAAACCCTAACCCCCGCTCCCGCCTCGTCCGCCTCCGAGCCGCATCGCGCGTG 68

Qy     185 TCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTC 244
           ||||||||| ||| | |||| | ||||||| ||||||| || |||| ||||| |||| |
Db      69 CCGTCCGCGGCCTTAAACGCGCGCGGGCGCCCGCCGATGTGTTCGGTCAGGGCTTCGTCC 128

Qy     245 GCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTG 304
           |||||||| |||| |||| || |||| ||||||||||||||||||||||||||| ||||||
Db     129 GCCGCTGGTGCGGGCGGCTGGGGAGCGGCGGCAGTGGGGGAGTTGGCGACGGAACGGCTG 188

Qy     305 GTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGC 364
           |||||||||||||| |||||||| ||||| || |||||||||||| ||||||||||||
Db     189 GTGGAGGTGGCGCAGCGGGCGGCAGACGCGGCCGGGGAGGTGCTCAGGAAGTACTTCCGC 248

Qy     365 CAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGA 424
           |||||||||||||||||||||| |||||| |||| |||||||| |||||||||| |||
Db     249 CAGCGGGTTGAGATCATCGACAAGGAGGACAACAGCCCTGTTACGATTGCAGATAG-AGA 307

Qy     425 AGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGG 484
           |||||||||||||||| |||||||||||||||||||||||||||| |||| ||||||
Db     308 AGCAGAAGAAGCAATGGTGTCAGTTATACTGAAGAGCTTCCCTACTCATGCCATTTTTGG 367

Qy     485 NGAAGAGA 492
           || ||||
Db     368 TGAGGAGA 375


RESULT 3
BQ240879
LOCUS       BQ240879                 444 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE05011G09R TaE05 Triticum aestivum cDNA clone TaE05011G09R, mRNA
            sequence.
ACCESSION   BQ240879
VERSION     BQ240879.1  GI:20436755
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 444)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >2.0 kb
            Plate: 011  row: G  column: 09
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
```

```
     source          1. .444
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE05011G09R"
                     /clone_lib="TaE05"
                     /tissue_type="developing seeds"
                     /dev_stage="5 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pSPORT-P (Invitrogen Technologies); Site_1:
                     NotI; Site_2: MluI; mRNA obtained from wheat seeds of
                     cultivar Glenlea 5 days post-anthesis"
BASE COUNT       76 a    141 c    146 g     81 t
ORIGIN

  Query Match             36.0%;  Score 177.2;  DB 14;  Length 444;
  Best Local Similarity   69.1%;  Pred. No. 5.5e-26;
  Matches  288;  Conservative    0;  Mismatches  116;  Indels   13;  Gaps    3;

Qy      88 CTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCG 147
           ||||||| | |  ||||||  ||| | |||   | ||||| |||||| |||| || |||
Db      25 CTCCACCATCCCACCCACCTTTCCCCTTCCCCATCCCGCCTTCGCAAAACCTAGCCTCCG 84

Qy     148 CTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTT--------- 198
           | ||||| |   | | |  |    |  ||  || |||||||
Db      85 TCACCACCTCCGCTCGCCCTCCCTCCTGGCAACCTTCTCCTCCGCGGCGGCAGGTCGGGC 144

Qy     199 GAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGC 258
           | ||| ||  | | |||    |||| |  ||||| | ||||||  | | |  | || ||
Db     145 GTGCGGGATAGCGGGCCGTTGGATGGGCTCGGTTCGAGCCTCGCCCTCTGAGGCGGAGGG 204

Qy     259 CGGCCGGAGAGCTGCGGCAG---TGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGC 315
           | |   |  |||||||| |    || || | ||| || ||||||||||||||| ||||||
Db     205 CTGGGCGGTGGCTGCGGCGGGTAAGGAGGGGGTGGAGATGGAGCGGCTGGTGGCGGTGGC 264

Qy     316 GCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTGA 375
           |||  | |||||||| || || |||||||||||||| |||||| ||| | ||||  | ||
Db     265 GCAGAGCGCGGCGGATGCGGCGGGGGAGGTGCTCAGGAAGTATTTCAGGCAGCGCTTCGA 324

Qy     376 GATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGAAG 435
           |||||||||||| |||||||||||||| || || || || ||||| |||||||||  |||
Db     325 GATCATCGACAAGGAGGACCACAGTCCGGTCACGATCGCTGATAG-AGAAGCAGAGCAAG 383

Qy     436 CAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
           |||||  |||||| |||||||| ||||| |||||||| ||   ||| || || ||||
Db     384 CAATGACGTCAGTCATACTGAATAGCTTTCCTACTCATGCTGTTTTCGGCGAGGAGA 440


RESULT 4
BI778441
LOCUS       BI778441                 588 bp    mRNA    linear   EST 23-JUL-2002
DEFINITION  EBro07_SQ003_F02_R root, 3 week, reduced light, cv Optic, EBro07
            Hordeum vulgare cDNA clone EBro07_SQ003_F02 5', mRNA sequence.
ACCESSION   BI778441
VERSION     BI778441.2  GI:21950120
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 588)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     On Sep 26, 2001 this sequence version replaced gi:15781333.
            Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk
            All sequence has a Phred quality score of 20 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1. .588
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro07_SQ003_F02"
                     /clone_lib="root, 3 week, reduced light, cv Optic, EBro07"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from shoot and root material of 3 week old
                     etiolated barley plants.  Developed as part of the barley
                     transcriptome resources of BBSRC/SEERAD funded cereal IGF
                     (Investigating Gene Function) project."
BASE COUNT       98 a    200 c    183 g    106 t      1 others
ORIGIN

  Query Match             35.8%;  Score 176;  DB 13;  Length 588;
  Best Local Similarity   66.7%;  Pred. No. 9.7e-26;
  Matches  318;  Conservative    0;  Mismatches  142;  Indels   17;  Gaps    4;

Qy      28 CATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTC 87
           |||||||||||  ||    |  |  ||||   |  | |  ||  |    || ||    |  |
Db     104 CATTTAAACCTCGCCGTCTGCCGCCGCTTTCCCCCCTCCCAGATGCTTCCCC----CAAC 159

Qy      88 CTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCG 147
           |||||||  | |  ||||||  ||| | |||  || ||||| | |||| |||| |  |||
Db     160 CTCCACCATCCCACCCACCTTTCCCCTTCCCCACCCCGCCTTCACAAAACCTAGCTTCCG 219

Qy     148 CTCTCGCCTCCT---------CCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTT 198
           | |||||          || |||      ||||||  | | | || | |
Db     220 TCACCACCTCCGCTCGTCCTCCCTCCTGGCAAGCGCCTCCTCTGCGGCGGCGGGTCGGGC 279

Qy     199 GAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGC 258
           | ||| ||  | | |||    || |||  ||||| | ||||||    | |  | || ||
Db     280 GTGCGGGATAGCGGGCCGTTGGACGAGCTCGGTTCGAGCCTCGCCTTCTGAGGCGGGGGG 339

Qy     259 CGGCCGGAGAGCTGCGGCAG---TGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGC 315
           | |   |  |||||||| |    || || | ||| || ||||||||||||| ||||||
Db     340 CTGGGCGGTGGCTGCGGCTGGTAAGGAGGGGGTGGAGATGGAGCGGCTGGTGGCGGTGGC 399

Qy     316 GCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTGA 375
           |||  | || ||||| || || |||||||||||||||||||||| | |||||  | ||
Db     400 GCAGAGCGCAGCGGATGCGGCGGGGGAGGTGCTCANGAAGTACTTCAGGCAGCGCTTCGA 459

Qy     376 GATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGAAG 435
           ||| |||||||| |||||||||||||| || || || || ||||| ||||||||||| |
Db     460 GATAATCGACAAGGAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAGAGG 518

Qy     436 CAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
           |||||  ||||| |||||||||||||| |||||||| ||   ||| || || ||||
Db     519 CAATGACTTCAGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGCGAGGAGA 575


RESULT 5
BI951067
LOCUS       BI951067                 863 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSME10024E07f Hordeum vulgare spike EST library HVcDNA0012
            (Fusarium infected) Hordeum vulgare cDNA clone HVSME10024E07f, mRNA
            sequence.
ACCESSION   BI951067
VERSION     BI951067.1  GI:16293708
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
```

```
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Wing,R., Muehlbauer,G.J., Close,T.J., Kleinhofs,A., Wise,R., Heinen
            ,S., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
            Simmons,J., Fenton,R.D., Malatrasi,M., Choi,D.W., Oates,R. and Main
            ,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Fusarium infected Morex spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 451
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 2
            High quality sequence stop: 575.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSME10024E07f"
                     /clone_lib="Hordeum vulgare spike EST library HVcDNA0012
                     (Fusarium infected)"
                     /tissue_type="Spike"
                     /lab_host="TJC121"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown at the University of Minnesota in
                     the GJ Muehlbauer lab; spikes were harvested and snap
                     frozen at 0, 1, 2, 3, 4, 5, 6, and 8 days after Fusarium
                     graminearum inoculation (Heinen). In the TJ Close lab at
                     the University of California, Riverside, total RNA was
                     prepared from each sample pool, equal quantities of all
                     eight RNA pools were combined, poly(A) RNA was purified
                     from the mixture, one primary unamplified cDNA library was
                     made, and 1 million pfu were in vivo excised to give
                     pBluescript SK(-) cDNA phagemids (Choi, Fenton, Malatrasi
                     ). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      184 a    254 c    240 g    183 t      2 others
ORIGIN

  Query Match             35.6%;  Score 175;  DB 13;  Length 863;
  Best Local Similarity   68.9%;  Pred. No. 1.6e-25;
  Matches  288;  Conservative    0;  Mismatches  117;  Indels   13;  Gaps    3;

Qy      87 CCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCC 146
           ||||||||  | |  ||||||  ||| | |||  || ||||| | ||||,|||| |  ||
Db      12 CCTCCACCATCCCACCCACCTTTCCCCTTCCCCACCCCGCCTTCACAAAACCTAGCTTCC 71

Qy     147 GCTCTCGCCTCCT---------CCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCT 197
           |    | |||||          || |||      ||||||  | | | || | |
Db      72 GTCACCACCTCCGCTCGTCCTCCCTCCTGGCAAGCGCCTCCTCTGCGGCGGCGGGTCGGG 131

Qy     198 TGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGG 257
           | ||| ||  | | |||    || |||  ||||| | ||||||    | |  | || ||
Db     132 CGTGCGGGATAGCGGGCCGTTGGACGAGCTCGGTTCGAGCCTCGCCTTCTGAGGCGGGGG 191

Qy     258 CCGGCCGGAGAGCTGCGGCAG---TGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGG 314
           | |   |  ||||||| |    || || | ||| || |||||||||||||| |||||
Db     192 GCTGGGCGGTGGCTGCGGCTGGTAAGGAGGGGGTGGAGATGGAGCGGCTGGTGGCGGTGG 251

Qy     315 CGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTG 374
           ||||  | || ||||| || || |||||||||||||||||||||||| | |||||  | |
Db     252 CGCAGAGCGCAGCGGATGCGGCGGGGGAGGTGCTCANGAAGTACTTCAGGCAGCGCTTCG 311

Qy     375 AGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGAA 434
           |||| |||||||| |||||||||||||| || || || || ||||| ||||||||||||
Db     312 AGATAATCGACAAGGAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAGAG 370

Qy     435 GCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
           ||||||   ||||| |||||||||||||| ||||||| ||   ||| || || ||||
Db     371 GCAATGACTTCAGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGCGAGGAGA 428


RESULT 6
BM370373
LOCUS       BM370373                 598 bp    mRNA    linear   EST 23-JUL-2002
DEFINITION  EBro08_SQ004_A09_R root, 3 week, drought-stressed, cv Optic, EBro08
            Hordeum vulgare cDNA clone EBro08_SQ004_A09 5', mRNA sequence.
ACCESSION   BM370373
VERSION     BM370373.2  GI:21951481
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     On Jan 10, 2002 this sequence version replaced gi:18113763.
            Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk
            All sequence has a Phred quality score of 20 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro08_SQ004_A09"
                     /clone_lib="root, 3 week, drought-stressed, cv Optic,
                     EBro08"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old drought stressed barley
                     plants.  Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded cereal IGF (Investigating
                     Gene Function) project."
BASE COUNT      103 a    209 c    176 g    110 t
ORIGIN

  Query Match             34.8%;  Score 171.2;  DB 13;  Length 598;
  Best Local Similarity   65.4%;  Pred. No. 8.9e-25;
  Matches  317;  Conservative    0;  Mismatches  151;  Indels   17;  Gaps    4;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | | ||  |  |||||||||||  ||   |  |  ||||   |  | |  ||  |    ||
Db     119 CGCGGCGACACCATTTAAACCTCGCCGTCTGCCGCCGCTTTCCCCCCTCCCAGATGCTTC 178
```

```
Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
           ||    |  ||||||||  | |  ||||||  ||| | |||  || |||||  | ||||
Db     179 CCC----CAACCTCCACCATCCCACCCACCTTTCCCCTTCCCCACCCCGCCTTCACAAAA 234

Qy     137 CCTAACCCCCGCTCTCGCCTCCT---------CCGCCTCCGCGCCGCCTCGCCCGTGTCG 187
           ||||  |  |||    | |||||          || |||      ||||||  | | | ||
Db     235 CCTAGCTTCCGTCACCACCTCCGCTCGTCCTCCCTCCTGGCAAGCGCCTCCTCTGCGGCG 294

Qy     188 TCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCC 247
           | |       | ||| ||  | | |||    || |||  |||||  | ||||||    |
Db     295 GCGGGTCGGGCGTGCGGGATAGCGGGCCGTTGGACGAGCTCGGTTCGAGCCTCGCCTTCT 354

Qy     248 GCTGGGGCGGCCGGCCGGAGAGCTGCGGCAG---TGGGGGAGTTGGCGACGGAGCGGCTG 304
           |  | || || | |   |  ||||||||  |    || || | ||| || ||||||||||
Db     355 GAGGCGGGGGGCTGGGCGGTGGCTGCGGCTGGTAAGGAGGGGGTGGAGATGGAGCGGCTG 414

Qy     305 GTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGC 364
           ||||  |||||||||  | ||  ||||  || || ||||||||||||  |||||||||  |
Db     415 GTGGCGGTGGCGCATAGCGCATCGGATGCGGCGGGGGAGGTGCTCATGAAGTACTTCAGG 474

Qy     365 CAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGA 424
           |||||  |  ||||  |||||||| |||||||||||||| || || || || |||||  |||
Db     475 CAGCGCTTCCAGATAATCGACAAGGAGGACCACAGTCCCGTCACGATCGCTGATAG-AGA 533

Qy     125 AGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGC TCCCTACTCAAGCCAATTTTGG 484
           |||||||||  ||||||   |||||  |||||||||||||  |||||||  ||   |||  ||
Db     534 AGCAGAAGAGGCAATGACTTCAGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGG 593

Qy     485 NGAAG 489
           ||  |
Db     594 CGAGG 598
```

RESULT 7
AU093884

LOCUS	AU093884	483 bp	mRNA	linear	EST 03-APR-2002

DEFINITION AU093884 Rice panicle at flowering stage Oryza sativa (japonica cultivar-group) cDNA clone E1081, mRNA sequence.
ACCESSION AU093884
VERSION AU093884.1 GI:8856566
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 483)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from panicle at flowering stage (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT ='RGP'.

FEATURES Location/Qualifiers
source 1. .483
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E1081"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 87 a 138 c 172 g 84 t 2 others
ORIGIN

Query Match 34.2%; Score 168.4; DB 9; Length 483;
Best Local Similarity 64.7%; Pred. No. 3.2e-24;
Matches 280; Conservative 0; Mismatches 146; Indels 7; Gaps 2;

```
Qy      66 CCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCG 125
           ||| |||||||  |   |  |||||  ||  ||   |  ||  |  |    ||  |  || ||
Db      18 CCCAAATGCTCCCGGCCACCTCCACCGTCCCACCCACCTCGCCCCTTCCCCACTCCCGCG 77

Qy     126 CCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGT 185
           |  || |   ||   ||| |||| |  | || |||| |     ||  ||  ||   | |
Db      78 CTGCCCCGCGCCGCCACCTCCGCGCAGGGCTTCTCCACGGGTCCGTGGCGACGGNGGCGG 137

Qy     186 CGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCG 245
           |     ||  |   | | ||    | |||| |   ||| | ||    ||    |
Db     138 TGGTTCGGGCGTGCGGGGTGACGACGGGCCACCGTCGGAGGATGGGCTCGGTTCGGGCGA 197

Qy     246 CCGCTGGGGCGGCCGC CGGAGAGCTGCGGCAGTGGG------GGAGTTGGCGACGGAGC 299
           | |  |||| | |||  ||   |  |||   | ||       || | | | |  |||||
Db     198 GCCCCAGGGCAGGCGGATGGGAGGTGGCGATGGCGGATAAGGAGGGGGTAGGGGTGGAGC 257

Qy     300 GGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACT 359
           ||||  |||||||||||||  |||||||  |||||  ||  |||||||||||||  ||||||||
Db     258 GGCTANTGGAGGTGGCGCAGAGGGCGGCAGACGCCGCAGGGGAGGTGCTCAGGAAGTACT 317

Qy     360 TCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATA 419
           ||||  ||  ||  ||  |||||||||||||  |||||  ||  ||||||||  ||  ||||||||||
Db     318 TCCGGCAACGCGTCGAGATCATCGACAAGGAGGATCAGAGTCCTGTAACTATTGCAGATA 377

Qy     420 GAAGAAGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAAT 479
           |  ||||||  ||||||||||||||  ||||||  ||  ||||||||||||||  | ||  |
Db     378 G-AGAAGCGGAAGAAGCAATGGTGTCAGTCATTCTGAAGAGCTTCCCTTCCCATGCCGTT 436

Qy     480 TTTGGNGAAGAGA 492
           |||||  ||  ||  |
Db     437 TTTGGTGAGGAAA 449
```

RESULT 8
BE418829

LOCUS	BE418829	864 bp	mRNA	linear	EST 24-JUL-2000

DEFINITION SCL081.H04R990808 ITEC SCL Wheat Leaf Library Triticum aestivum cDNA clone SCL081.H04, mRNA sequence.
ACCESSION BE418829
VERSION BE418829.1 GI:9416675
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 864)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.
TITLE International Triticeae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae
JOURNAL Unpublished (2000)
COMMENT Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.

FEATURES Location/Qualifiers
source 1. .864
/organism="Triticum aestivum"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="SCL081.H04"

/clone_lib="ITEC SCL Wheat Leaf Library"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."
BASE COUNT 194 a 243 c 242 g 171 t 14 others
ORIGIN

Query Match 34.2%; Score 168.4; DB 10; Length 864;
Best Local Similarity 67.3%; Pred. No. 3.3e-24;
Matches 329; Conservative 0; Mismatches 141; Indels 19; Gaps 6;

```
Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | | ||  |  |||||||||||| ||  ||  |  ||||       |||||||  |||||
Db     115 CGCGGCGACACCATTTAAACCTCGCCGCCTGCCGCCGC--TTTTTTCCTCCCAGATGCTT 172

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
            ||  |   ||  | |||  ||   ||  |||||  ||| | |||  || ||||| | ||||
Db     173 CCCCCAATCTGCACCATCC---CGCTCACCTTTCCCCTTCCCAACCCCGCCTTCACAAAA 229

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||||  ||  |||    | |||||  |   | |  |  |    |  || | ||  |||||||
Db     230 CCTAGCCTCCGTCACCACCTCCGCTCGCCCTCCCTCCTGGCAACCTTCTCCTCCGCGGCG 289

Qy     197 TTGAGC---------GCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCC 247
             | |          | ||  | | |||    ||||  |  |||||  | ||||||  |  |
Db     290 GCGGGTCGGGCGTGTGGGATAGCGGGCCGTTGGATGGGCTCGGTTCGAGCCTCGCCCTCT 349

Qy     248 GCTGGGGCGGCCGGCCGGAGAGCTGCGGCAG---TGGGGGAGTTGGCGACGGAGCGGCTG 304
           |  | || || | |   |  |||||||| |    ||  | | ||| || ||||||||||
Db     350 GAGGCGGGGGGCTGGGCGGTGGCTGCGGCGGGTAAGGAAGGGGTGGAGATGGAGCGGCTG 409

Qy     305 GTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGC 364
           |||| ||||||  ||  | ||||||||  || || ||||||||||||  |||||||||| |
Db     410 GTGGCGGTGGCCCANAGCGCGGCGGATGCGGCGGGGG\GGTGCTCAGGAAGTACTTCAGG 469

Qy     365 CAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGA 424
           |||||  | ||||||||||||||||||||||||||| || || || || ||||| |||
Db     470 CAGCGCTTCGAGATCATCGACAAAGAGGACCACAGTCCGGTCACGATCGCTGATAG-AGA 528

Qy     425 AGCAGAAGAAGCAATGGNGTCAGTTATACTGAAG-AGCTTCCCTACTCAAGCCAATTTTG 483
           ||||||  |||||||||  ||||||  |||||||| ||||| |||||||| ||   ||| |
Db     529 AGCAGAGGAAGCAATGACGTCAGTCATACTGAAGAAGCTTTCCTACTCATGCTGTTTTCG 588

Qy     484 GNGAAGAGA 492
           | |||||  |
Db     589 GCGAAGANA 597
```

ESULT 9
F631396
LOCUS BF631396 766 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEb0015N03f Hordeum vulgare seedling shoot EST library HVcDNA0002 (Dehydration stress) Hordeum vulgare cDNA clone HVSMEb0015N03f, mRNA sequence.
ACCESSION BF631396
VERSION BF631396.2 GI:13092082
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 766)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex drought-stressed seedling shoot cDNA library
JOURNAL Unpublished (2001)
COMMENT On Dec 19, 2000 this sequence version replaced gi:11895554.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 565
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 740.
FEATURES Location/Qualifiers
source 1. .766
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEb0015N03f"
/clone_lib="Hordeum vulgare seedling shoot EST library HVcDNA0002 (Dehydration stress)"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were incubated at 90% RH for 24 hr. Shoots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, 600000 pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 165 a 177 c 242 g 182 t
ORIGIN

Query Match 32.7%; Score 161; DB 12; Length 766;
Best Local Similarity 68.9%; Pred. No. 1e-22;
Matches 248; Conservative 0; Mismatches 108; Indels 4; Gaps 2;

```
Qy     136 CCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGT 195
           || | ||| || |        |  || |||      ||||||  | | | || | |
Db      30 CCGTCACCACCTCCGCTCGTCCTCCCTCCTGGCAAGCGCCTCCTCTGCGGCGGCGGGTCG 89

Qy     196 CTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGC 255
             | ||| ||  | | |||    || |||  ||||| | ||||||    | |  | ||
Db      90 GGCGTGCGGGATAGCGGGCCGTTGGACGAGCTCGGTTCGAGCCTCGCCTTCTGAGGCGGG 149

Qy     256 GGCCGGCCGGAGAGCTGCGGCAGT---GGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGT 312
           || | |   |   |||||||| |    || || | ||| || ||||||||||||| |||
Db     150 GGGCTGGGCGGTGGCTGCGGCTGGTAAGGAGGGGGTGGAGATGGAGCGGCTGGTGGCGGT 209

Qy     313 GGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGT 372
           ||||||  | || |||| || || |||||||||||| |||||||||| | |||||  |
Db     210 GGCGCAGAGCGCAGCGGATGCGGCGGGGGAGGTGCTCAGGAAGTACTTCAGGCAGCGCTT 269

Qy     373 TGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAG 432
           |||||  |||||||| |||||||||||||| || || || || ||||| ||||||||||
Db     270 CGAGATAATCGACAAGGAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAG 328

Qy     433 AAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
           | ||||||   ||||| |||||||||||||| |||||||| ||   ||| || || ||||
```

```
Db     329 AGGCAATGACTTCAGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGCGAGGAGA 388

RESULT 10
BQ467759
LOCUS       BQ467759                443 bp    mRNA    linear   EST 30-MAY-2002
DEFINITION  HS04M21r HS Hordeum vulgare cDNA clone HS04M21 5-PRIME, mRNA
            sequence.
ACCESSION   BQ467759
VERSION     BQ467759.1  GI:21275541
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner
            ,A.
  TITLE     Barley ESTs from germinating seeds
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Corrensstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 443    Std Error: 0.00
            Plate: 4  row: M  column: 21
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Hordeum vulgare"
                     /cultivar="Barke"
                     /db_xref="taxon:4513"
                     /clone="HS04M21"
                     /clone_lib="HS"
                     /tissue_type="embryo + scutellum"
                     /dev_stage="0-16 hours after imbibition"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
                     artefact caused by the kit, in most cases the EcoRI site
                     is NOT present, as well as the EcoRIadapter used for
                     cloning. To excise the insert, restriction sites upstream
                     EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also
                     due to the cloning system used Blue/white selection for
                     recombinats is not 100% reliable."
BASE COUNT       93 a     94 c    161 g     95 t
ORIGIN

  Query Match           31.5%;  Score 155.2;  DB 14;  Length 443;
  Best Local Similarity  74.6%;  Pred. No. 1.4e-21;
  Matches  220;  Conservative    0;  Mismatches   71;  Indels    4;  Gaps    2;

Qy     201 GCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCG 260
           || |||| | | |||    || |||  ||||| | ||||||    | |  | || || |
Db       3 GCACGAGGGCGGGCCGTTGGACGAGCTCGGTTCGAGCCTCGCCTTCTGAGGCGGGGGGCT 62

Qy     261 GCCGGAGAGCTGCGGCAG---TGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGC 317
           |   |   |||||||| |    || || | ||| || |||||||||||||| ||||||||
Db      63 GGGCGGTGGCTGCGGCTGGTAAGGAGGGGGTGGAGATGGAGCGGCTGGTGGCGGTGGCGC 122

Qy     318 AACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTGAGA 377
           |  | || ||||| || || |||||||||||||| |||||||||| | |||||  | ||||
Db     123 AGAGCGCAGCGGATGCGGCGGGGGAGGTGCTCAGGAAGTACTTCAGGCAGCGCTTCGAGA 182

Qy     378 TCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGAAGCA 437
           | |||||||| |||||||||||||| || || || || ||||| |||||||||||| |||
Db     183 TAATCGACAAGGAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAGAGGCA 241

Qy     438 ATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492

           ||| |||||  |||||||||||||| |||||||| ||   ||| || || ||||
Db     242 ATGACTTCAGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGCGAGGAGA 296

RESULT 11
BE498543
LOCUS       BE498543                404 bp    mRNA    linear   EST 04-AUG-2000
DEFINITION  WHE0971_F05_K09ZS Wheat pre-anthesis spike cDNA library Triticum
            aestivum cDNA clone WHE0971_F05_K09, mRNA sequence.
ACCESSION   BE498543
VERSION     BE498543.1  GI:9697160
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 404)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0971_F05_K09"
                     /clone_lib="Wheat pre-anthesis spike cDNA library"
                     /tissue_type="Spike before anthesis"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse.  Whole spike with awns trimmed,  white, green
                     and yellow anther were collected and total RNA, and
                     poly(A) RNA were prepared, a cDNA library was made, and
                     the cDNA  clones were in vivo excised to give pBluescript
                     phagemids in the TJ  Close lab (Choi, Close, Fenton) at
                     the University of California,  Riverside.  Plasmid DNA
                     preparations and DNA sequencing were performed  in the OD
                     Anderson lab (all other authors)."
BASE COUNT       91 a     83 c    138 g     89 t      3 others
ORIGIN

  Query Match           30.2%;  Score 148.8;  DB 10;  Length 404;
  Best Local Similarity  81.7%;  Pred. No. 2.8e-20;
  Matches  183;  Conservative    0;  Mismatches   40;  Indels    1;  Gaps    1;

Qy     269 GCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCG 328
           || ||||    || || | |||  | |||||||||||||| |||||||||  | ||||||
Db      21 GCGGCGGGTAAGGAGGGGGTGGACATGGAGCGGCTGGTGGCGGTGGCGCANAGCGCGGCG 80

Qy     329 GACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTGAGATCATCGACAAA 388
           || || || |||||||||||||||||||||||| | |||||  | ||||||||||||||
Db      81 GATGCGGCGGGGGAGGTGCTCANGAAGTACTTCAGGCAGCGCTTCGAGATCATCGACAAA 140

Qy     389 GAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGAAGCAATGGNGTCAGT 448
           |||||||||||||| || || || ||||| ||||||||||||||||||   |||||
Db     141 GAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAGAAGCAATGACCTCAGT 199
```

```
Qy     449 TATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
           ||||||||||||| |||||||| ||   ||| || || ||||
Db     200 CATACTCAAGAGCTTTCCTACTCATGCTGTTTTCGGTGAGGAGA 243
```

RESULT 12
BI959675
LOCUS BI959675 667 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEn0020L21f Hordeum vulgare rachis EST library HVcDNA0015 (normal) Hordeum vulgare cDNA clone HVSMEn0020L21f, mRNA sequence.
ACCESSION BI959675
VERSION BI959675.1 GI:16310930
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 667)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons ,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex rachis cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 339
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 542.

FEATURES Location/Qualifiers
source 1. .667
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEn0020L21f"
/clone_lib="Hordeum vulgare rachis EST library HVcDNA0015 (normal)"
/tissue_type="Rachis"
/lab_host="TJC121"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plants were grown at Washington State University, Pullman, WA in a greenhouse, the rachises were excised and frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close lab at the University of California, Riverside total RNA was prepared, poly(A) was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 119 a 222 c 198 g 113 t 15 others
ORIGIN

Query Match 29.5%; Score 145; DB 13; Length 667;
Best Local Similarity 62.8%; Pred. No. 1.7e-19;

Matches 267; Conservative 0; Mismatches 145; Indels 13; Gaps 3;

```
Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
           || || |   | |     |  | ||||||||  | |  ||||||  ||| | |||  ||
Db      38 CCCCCTCCCAGATGCTTCCCCCAACCTCCACCATCCCACCCACCTTTCCCCTTCCCCACC 97

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCT---------CCGCCTCCGCGCCG 173
           ||||| | |||| |||| |  |||     | |||||          || |||      ||
Db      98 CCGCCTTCACAAAACCTAGCTTCCGTCACCACCTCCGCTCGTCCTCCCTCCTGGCAAGCG 157

Qy     174 CCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTA 233
           ||||  | | | || | |       | ||| ||  | | |||    || |||  |||||
Db     158 CCTCCTCTGCGGCGGCGGGTCGGGCGTGCGGGATAGCGGGCCGTTGGACGAGCTCGGTTC 217

Qy     234 GGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAG---TGGGGGAGTTGG 290
           | ||||||    | |  | || || | |   |  ||||||||  |    || || | |||
Db     218 GAGCCTCGCCTTCTGAGGCGGGGGGCTGGGCGGTGGCTGCGGCTGGTAAGGAGGGGGTGG 277

Qy     291 CGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCA 350
           || ||| |||||||||| || |||||     ||  |||| || || ||||||| |||||
Db     278 AGATGGANCGGCTGGTGGCGGNGGCGCNNANCGCANCGGATGCGGCGGGGGAGGGGCTCA 337

Qy     351 NGAAGTACTTCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAA 410
           |||||||||    ||||| | ||||| |||||||| ||||||||||||  ||    |
Db     338 NGAAGTACTTGAAGCAGCGCTTCGAGATAATCGACAAGGAGGACCACAGTCGCGTACGGA 397

Qy     411 TTGCAGATAGAAGAAGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACT 470
           |   |   | |||||| | || || | ||  |||| ||||||||||||  ||||||
Db     398 NNGNNGNGNG-AGAAGCAAANGAGGCGAAGGACTCAGTCATACTGAAGAGCTTNCCTACT 456

Qy     471 CAAGC 475
           || ||
Db     457 CATGC 461
```

RESULT 13
BG414965
LOCUS BG414965 631 bp mRNA linear EST 23-OCT-2001
DEFINITION HVSMEk0004H12f Hordeum vulgare testa/pericarp EST library HVcDNA0013 (normal) Hordeum vulgare cDNA clone HVSMEk0004H12f, mRNA sequence.
ACCESSION BG414965
VERSION BG414965.1 GI:13320516
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 631)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Kannangara,G., von Wetstein,D., Akhunov,E., Chin,A., Choi,D.W., Fenton,R.D., Kianian ,P., Otto,C., Simons,K., Zhang,D., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex testa/pericarp cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 255
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 439.

FEATURES Location/Qualifiers
source 1. .631
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEk0004H12f"

/clone_lib="Hordeum vulgare testa/pericarp EST library HVcDNA0013 (normal)"
/tissue_type="testa/pericarp"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Plants were raised from seeds in a Controlled Environments growth chamber maintained in continuous light at 18oC, and testa and pericarp were dissected from developing kernels at Washington State University, Pullman, WA (Kannangara, von Wetstein). Total RNA was prepared, poly(A) RNA was purified, one cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids in the TJ Close lab at the University of California, Riverside (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 153 a 127 c 193 g 157 t 1 others
ORIGIN

Query Match 28.8%; Score 141.8; DB 12; Length 631;
Best Local Similarity 79.5%; Pred. No. 7.3e-19;
Matches 178; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

```
Qy    269 GCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCG 328
          || || |    || || | ||| || |||||||||||||| ||||||||  | || |||
Db      1 GCGGCTGGTAAGGAGGGGGTGGAGATGGAGCGGCTGGTGGCGGTGGCGCAGAGCGCAGCG 60

Qy    329 GACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTGAGATCATCGACAAA 388
          || || || |||||||||||||| |||||||||| | ||||| | ||||| ||||||||
Db     61 GATGCGGCGGGGGAGGTGCTCAGGAAGTACTTCAGGCAGCGCTTCGAGATAATCGACAAG 120

Qy    389 GAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGAAGCAATGGNGTCAGT 448
          |||||||||||||| || || || || ||||| |||||||||||| ||||||   |||||
Db    121 GAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAGAGGCAATGACTTCAGT 179

Qy    449 TATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
          |||||||||||||| |||||||| ||   ||| || || ||||
Db    180 CATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGCGAGGAGA 223
```

RESULT 14
BQ801093
LOCUS BQ801093 759 bp mRNA linear EST 30-JUL-2002
DEFINITION WHE2810_C03_E06ZS Triticum monococcum vernalized apex cDNA library Triticum monococcum cDNA clone WHE2810_C03_E06, mRNA sequence.
ACCESSION BQ801093
VERSION BQ801093.1 GI:22016062
KEYWORDS EST.
SOURCE Triticum monococcum.
ORGANISM Triticum monococcum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 759)
AUTHORS Anderson,O.D., Chao,S., Crossman,C., Dubcovsky,J., Echenique,V., Lazo,G.R., Pham,J., Rausch,C.J., Stamova,B., Wilson,C. and Woo,J.
TITLE The structure and function of the expressed portion of the wheat genomes - Vernalized apex cDNA library from Triticum monococcum
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES Location/Qualifiers
source 1. .759
/organism="Triticum monococcum"
/cultivar="G3116"
/db_xref="taxon:4568"
/clone="WHE2810_C03_E06"
/clone_lib="Triticum monococcum vernalized apex cDNA library"
/tissue_type="Vernalized apex"
/dev_stage="One month old plants"
/lab_host="E. coli XLOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; One-month old plants were subjected to vernalization treatment by placing them in the cold room at 6 C, under 15hr light/9hr dark condition. Total RNA was prepared from apex tissue extracted from plants with no cold treatment; and from plants with 2-week , 4-week and 6-week cold treatment separately. Equal amount of total RNA was pooled from all four samples, a cDNA library was made using pooled polyA RNA and cDNA clones were in vivo excised at the University of California, Davis (V. Echenique, B. Stamova, J. Dubcovsky). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT 186 a 153 c 223 g 197 t
ORIGIN

Query Match 24.2%; Score 119.2; DB 14; Length 759;
Best Local Similarity 82.1%; Pred. No. 2.6e-14;
Matches 147; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

```
Qy    314 GCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTT 373
          | | |  | |||||||| || || |||||||||||| |||||||||| | |||||  |
Db     10 GAGGAGAGCGCGGCGGATGCGGCGGGGGAGGTGCTCAGGAAGTACTTCAGGCAGCGCTTC 69

Qy    374 GAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGA 433
          |||||||||||||||||||||||||||||| || || || || ||||| |||||||| ||
Db     70 GAGATCATCGACAAAGAGGACCACAGTCCTGGTCACGATCGCTGATAG-AGAAGCAGAGGA 128

Qy    434 AGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
          |||||||  |||||| |||||||||||||| |||||||| ||   ||| || || ||||
Db    129 AGCAATGACGTCAGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGCGAGGAGA 187
```

RESULT 15
BQ608457
LOCUS BQ608457 745 bp mRNA linear EST 25-JUN-2002
DEFINITION BRY_4362 wheat EST endosperm library Triticum aestivum cDNA 5', mRNA sequence
ACCESSION BQ608457
VERSION BQ608457.1 GI:21557796
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 745)
AUTHORS Clarke,B., Lambrecht,M. and Rhee,S.
TITLE Assessing the utility of Arabidopsis genomic information for interpreting wheat EST sequences
JOURNAL Unpublished (2002)
COMMENT Contact: Lambrecht M
The Arabidopsis Information Resource

Carnegie Institution of Washington, Dept. of Plant Biology
260 Panama Street, Stanford, CA 94305, USA
Tel: 1 650 325 1521 x 251
Fax: 1 650 325 3748
Email: rhee@acoma.stanford.edu.

```
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Triticum aestivum"
                     /cultivar="Wyuna"
                     /db_xref="taxon:4565"
                     /clone_lib="wheat EST endosperm library"
                     /tissue_type="endosperm"
                     /dev_stage="developing endosperm tissue 8, 10 and 12 DPA
                     (days post anthesis)"
                     /note="Vector: Bluescript II SK(-)"
BASE COUNT      191 a    137 c    217 g    195 t      5 others
ORIGIN

  Query Match          23.4%;  Score 115.2;  DB 14;  Length 745;
  Best Local Similarity   83.8%;  Pred. No. 1.6e-13;
  Matches  140;  Conservative    0;  Mismatches   26;  Indels    1;  Gaps    1;

Qy     326 GCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCNCCAGCGGGTTGAGATCATCGAC 385
           ||||| || || |||||||||||||| |||||||||| | |||||  | ||||||||||||
Db       1 GCGGATGCGGCGGGGGAGGTGCTCAAGAAGTACTTCAGGCAGCGCTTCGAGATCATCGAC 60

Qy     386 AAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAGAAGCAATGGNGTC 445
           |||||||||||||||||| || || || || ||||| ||||||||||||||||||||   ||
Db      61 AAAGAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAGAAGCAATGACCTC 119

Qy     446 AGTTATACTGAAGAGCTTCCCTACTCAAGCCAATTTTGGNGAAGAGA 492
           ||| |||||||||||||| |||||||| ||   ||| || || ||||
Db     120 AGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGTGAGGAGA 166


Search completed: November 18, 2002, 20:55:28
Job time : 2122 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 19, 2002, 00:20:30 ; Search time 2718 Seconds
(without alignments)
5268.058 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agctccgagcgtcattctcc.....agccaattttggngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	112	22.8	131704	2	AP004313	AP004313 Oryza sat
	2	112	22.8	161809	2	AP005173	AP005173 Oryza sat
	3	84.2	17.1	163381	2	AC128230	AC128230 Rattus no
	4	78	15.9	177168	2	AC107145	AC107145 Rattus no
	5	76.8	15.6	138467	2	AC111697	AC111697 Rattus no
	6	74.4	15.1	82815	2	AC062001	AC062001 Homo sapi
	7	73.2	14.9	158897	2	AC121439	AC121439 Rattus no
	8	72.2	14.7	68330	2	AC116109	AC116109 Mus muscu
	9	72.2	14.7	117082	2	AC126076	AC126076 Rattus no
c	10	70.4	14.3	101509	2	AC027353	AC027353 Homo sapi
	11	69.8	14.2	151358	2	AC125906	AC125906 Rattus no
	12	68.4	13.9	174605	2	AC121227	AC121227 Rattus no
	13	68	13.8	174031	2	AC118303	AC118303 Rattus no
	14	67.4	13.7	68787	2	AC106398	AC106398 Rattus no
	15	67.4	13.7	130899	2	AC126730	AC126730 Rattus no
	16	67.4	13.7	165988	2	AC121212	AC121212 Rattus no
	17	67.2	13.7	185263	2	AC129765	AC129765 Rattus no
	18	67.2	13.7	188231	2	AC094753	AC094753 Rattus no
	19	66.8	13.6	179608	2	AC128497	AC128497 Rattus no
	20	66.8	13.6	209216	2	AC117126	AC117126 Rattus no
c	21	66.6	13.5	99849	2	AC096824	AC096824 Rattus no
c	22	66.6	13.5	117082	2	AC126076	AC126076 Rattus no
c	23	66.6	13.5	118276	2	AP004120	AP004120 Oryza sat
	24	66.6	13.5	158117	2	AP005111	AP005111 Oryza sat
c	25	66.6	13.5	158971	2	AP004882	AP004882 Oryza sat
c	26	66.6	13.5	165988	2	AC121212	AC121212 Rattus no
	27	66.6	13.5	171624	2	AC095621	AC095621 Rattus no
	28	66.2	13.5	194776	2	AC129706	AC129706 Rattus no
	29	65.8	13.4	171204	2	AC130936	AC130936 Rattus no
	30	65.6	13.3	175041	2	AC097133	AC097133 Rattus no
	31	65.2	13.3	62839	2	AC098999	AC098999 Rattus no
	32	65.2	13.3	170696	2	AP004658	AP004658 Oryza sat
c	33	65.2	13.3	171204	2	AC130936	AC130936 Rattus no
c	34	65	13.2	184245	2	AC127355	AC127355 Mus muscu
	35	64.6	13.1	110000	2	AC116301_0	AC116301 Homo sapi
	36	64.4	13.1	171852	2	AC058782	AC058782 Homo sapi
	37	64.4	13.1	181022	2	AC121905	AC121905 Mus muscu
	38	64.4	13.1	188133	2	AC112428	AC112428 Rattus no
	39	64.2	13.0	81417	2	AC023265	AC023265 Homo sapi
	40	64	13.0	69417	2	AC121516	AC121516 Mus muscu
	41	64	13.0	162183	2	AC097753	AC097753 Rattus no
	42	63.8	13.0	161163	2	AC115099	AC115099 Homo sapi
	43	63.8	13.0	178941	2	AC113635	AC113635 Rattus no
c	44	63.6	12.9	129667	2	AC095300	AC095300 Rattus no
c	45	63.6	12.9	135119	2	AC011578	AC011578 Homo sapi

ALIGNMENTS

RESULT 1
AP004313/c
LOCUS AP004313 131704 bp DNA linear HTG 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 7 clone P0440B02,
*** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AP004313
VERSION AP004313.1 GI:16874543
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:P0440B02.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1

```
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
            clone:P0440B02
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 131704)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .131704
                     /organism="Oryza sativa (japonica cultivar-group "
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="7"
                     /clone="P0440B02"
BASE COUNT    38468 a  27101 c  28237 g  37798 t    100 others
ORIGIN

  Query Match             22.8%;  Score 112;  DB 2;  Length 131704;
  Best Local Similarity   60.1%;  Pred. No. 4.5e-11;
  Matches  206;  Conservative    0;  Mismatches  131;  Indels    6;  Gaps    1;

Qy      66 CCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCG 125
           ||| ||||||| |  |  ||||| ||  ||   |  || |  | |    || |  || ||
Db 108777 CCCAAATGCTCCCGGCCACCTCCACCGTCCCACCCACCTCGCCCCTTCCCCACTCCCGCG 108718

Qy     126 CCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGT 185
           |  || |   ||   ||| |||| |  | || |||| |     ||  ||  || | | |
Db 108717 CTGCCCCGCGCCGCCACCTCCGCGCAGGGCTTCTCCACGGGTCCGTGGCGACGGCGGCGG 108658

Qy     186 CGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCG 245
           |     ||  |   | | ||    | |||| |    ||| | ||    ||    |
Db 108657 TGGTTCGGGCGTGCGGGGTGACGACGGGCCACCGTCGGAGGATGGGCTCGGTTCGGGCGA 108598

Qy     246 CCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGG------GGAGTTGGCGACGGAGC 299
           | |  |||| | |||  ||   |  |||   | ||       || | | | |  |||||
Db 108597 GCCCCAGGGCAGGCGGATGGGAGGTGGCGATGGCGGATAAGGAGGGGGTAGGGGTGGAGC 108538

Qy     300 GGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACT 359
           |||| ||||||||||||||  ||||||| ||||| || ||||||||||||| ||||||||
Db 108537 GGCTAGTGGAGGTGGCGCAGAGGGCGGCAGACGCCGCAGGGGAGGTGCTCAGGAAGTACT 108478

Qy     360 TCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCC 402
           |||| || || || |||||||||||||| ||||| || ||| |
Db 108477 TCCGGCAACGCGTCGAGATCATCGACAAGGAGGATCAGAGTGC 108435


RESULT 2
AP005173
LOCUS       AP005173               161809 bp    DNA     linear   HTG 17-MAY-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 7 clone
            OSJNBb0003E08, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP005173
VERSION     AP005173.1  GI:20975299
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OSJNBb0003E08.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OSJNBb0003E08
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 161809)
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .161809
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="7"
                     /clone="OSJNBb0003E08"
BASE COUNT    46758 a  34068 c  34038 g  46734 t    211 others
ORIGIN

  Query Match             22.8%;  Score 112;  DB 2;  Length 161809;
  Best Local Similarity   60.1%;  Pred. No. 4.4e-11;
  Matches  206;  Conservative    0;  Mismatches  131;  Indels    6;  Gaps    1;

Qy      66 CCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCG 125
           ||| ||||||| |  |  ||||| ||  ||   |  || |  | |    || |  || ||
Db 108667 CCCAAATGCTCCCGGCCACCTCCACCGTCCCACCCACCTCGCCCCTTCCCCACTCCCGCG 108726

Qy     126 CCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGT 185
           |  || |   ||   ||| |||| |  | || |||| |     ||  ||  || | | |
Db 108727 CTGCCCCGCGCCGCCACCTCCGCGCAGGGCTTCTCCACGGGTCCGTGGCGACGGCGGCGG 108786

Qy     186 CGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCG 245
           |     ||  |   | | ||    | |||| |    ||| | ||    ||    |
Db 108787 TGGTTCGGGCGTGCGGGGTGACGACGGGCCACCGTCGGAGGATGGGCTCGGTTCGGGCGA 108846

Qy     246 CCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGG------GGAGTTGGCGACGGAGC 299
           | |  |||| | |||  ||   |  |||   | ||       || | | | |  |||||
Db 108847 GCCCCAGGGCAGGCGGATGGGAGGTGGCGATGGCGGATAAGGAGGGGGTAGGGGTGGAGC 108906

Qy     300 GGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACT 359
           |||| ||||||||||||||  ||||||| ||||| || ||||||||||||| ||||||||
Db 108907 GGCTAGTGGAGGTGGCGCAGAGGGCGGCAGACGCCGCAGGGGAGGTGCTCAGGAAGTACT 108966

Qy     360 TCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCC 402
           |||| || || || |||||||||||||| ||||| || ||| |
Db 108967 TCCGGCAACGCGTCGAGATCATCGACAAGGAGGATCAGAGTGC 109009


RESULT 3
AC128230
LOCUS       AC128230               163381 bp    DNA     linear   HTG 19-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-57K8, *** SEQUENCING IN PROGRESS ***,
            85 unordered pieces.
ACCESSION   AC128230
VERSION     AC128230.1  GI:21908840
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
```

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 163381)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen  ., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfr d,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163381)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYBT
                Center clone name: CH230-57K8
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 78315 bases at least Q40
                Consensus quality: 87206 bases at least Q30
                Consensus quality: 92805 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 85 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1027: contig of 1027 bp in length
            *     1028     1127: gap of unknown length
            *     1128     2195: contig of 1068 bp in length
            *     2196     2295: gap of unknown length
            *     2296     3623: contig of 1328 bp in length
            *     3624     3723: gap of unknown length
            *     3724     5144: contig of 1421 bp in length
            *     5145     5244: gap of unknown length
            *     5245     6262: contig of 1018 bp in length
            *     6263     6362: gap of unknown length
            *     6363     7886: contig of 1524 bp in length
            *     7887     7986: gap of unknown length
            *     7987     9581: contig of 1595 bp in length
            *     9582     9681: gap of unknown length
            *     9682    11192: contig of 1511 bp in length
            *    11193    11292: gap of unknown length
            *    11293    12398: contig of 1106 bp in length
            *    12399    12498: gap of unknown length
            *    12499    13696: contig of 1198 bp in length
            *    13697    13796: gap of unknown length
            *    13797    14976: contig of 1180 bp in length
            *    14977    15076: gap of unknown length
            *    15077    16562: contig of 1486 bp in length
            *    16563    16662: gap of unknown length
            *    16663    17680: contig of 1018 bp in length
            *    17681    17780: gap of unknown length
            *    17781    18800: contig of 1020 bp in length
            *    18801    18900: gap of unknown length
            *    18901    20003: contig of 1103 bp in length
            *    20004    20103: gap of unknown length
            *    20104    21681: contig of 1578 bp in length
            *    21682    21781: gap of unknown length
            *    21782    22870: contig of 1089 bp in length
            *    22871    22970: gap of unknown length
            *    22971    24090: contig of 1120 bp in length
            *    24091    24190: gap of unknown length
            *    24191    25426: contig of 1236 bp in length
            *    25427    25526: gap of unknown length
            *    25527    26689: contig of 1163 bp in length
            *    26690    26789: gap of unknown length
            *    26790    28356: contig of 1567 bp in length
            *    28357    28456: gap of unknown length
            *    28457    30062: contig of 1606 bp in length
            *    30063    30162: gap of unknown length
            *    30163    31306: contig of 1144 bp in length
            *    31307    31406: gap of unknown length
            *    31407    32968: contig of 1562 bp in length
            *    32969    33068: gap of unknown length
            *    33069    34291: contig of 1223 bp in length
            *    34292    34391: gap of unknown length
            *    34392    35467: contig of 1076 bp in length
            *    35468    35567: gap of unknown length
            *    35568    36788: contig of 1221 bp in length
            *    36789    36888: gap of unknown length
            *    36889    38584: contig of 1696 bp in length
            *    38585    38684: gap of unknown length
            *    38685    40070: contig of 1386 bp in length
            *    40071    40170: gap of unknown length
            *    40171    41693: contig of 1523 bp in length
            *    41694    41793: gap of unknown length
            *    41794    42956: contig of 1163 bp in length
            *    42957    43056: gap of unknown length
            *    43057    44652: contig of 1596 bp in length
            *    44653    44752: gap of unknown length
            *    44753    46319: contig of 1567 bp in length
            *    46320    46419: gap of unknown length
            *    46420    48471: contig of 2052 bp in length
```

```
*    48472    48571: gap of unknown le gth
*    48572    49973: contig of 1402 bp in length
*    49974    50073: gap of unknown length
*    50074    51893: contig of 1820 bp in length
*    51894    51993: gap of unknown length
*    51994    53016: contig of 1023 bp in length
*    53017    53116: gap of unknown length
*    53117    54906: contig of 1790 bp in length
*    54907    55006: gap of unknown length
*    55007    56794: contig of 1788 bp in length
*    56795    56894: gap of unknown length
*    56895    58683: contig of 1789 bp in length
*    58684    58783: gap of unknown length
*    58784    60103: contig of 1320 bp in length
*    60104    60203: gap of unknown length
*    60204    61477: contig of 1274 bp in length
*    61478    61577: gap of unknown length
*    61578    62977: contig of 1400 bp in length
*    62978    63077: gap of unknown length
*    63078    64770: contig of 1693 bp in length
*    64771    64870: gap of unknown length
*    64871    66393: contig of 1523 bp in length
*    66394    66493: gap of unknown length
*    66494    67700: contig of 1207 bp in length
*    67701    67800: gap of unknown length
*    67801    68946: contig of 1146 bp in length
*    68947    69046: gap of unknown length
*    69047    70757: contig of 1711 bp in length
*    70758    70857: gap of unknown length
*    70858    72056: contig of 1199 bp in length
*    72057    72156: gap of unknown length
*    72157    73498: contig of 1342 bp in length
*    73499    73598: gap of unknown length
*    73599    75005: contig of 1407 bp in length
*    75006    75105: gap of unknown length
*    75106    76967: contig of 1862 bp in length
*    76968    77067: gap of unknown length
*    77068    78834: contig of 1767 bp in length
*    78835    78934: gap of unknown length
*    78935    80786: contig of 1852 bp in length
*    80787    80886: gap of unknown length
*    80887    82169: contig of 1283 bp in length
*    82170    82269: gap of unknown length
*    82270    83809: contig of 1540 bp in length
*    83810    83909: gap of unknown length
*    83910    85248: contig of 1339 bp in length
*    85249    85348: gap of unknown length

  Query Match             17.1%;  Score 84.2;  DB 2;  Length 163381;
  Best Local Similarity   52.3%;  Pred. No. 5.6e-06;
  Matches  172;  Conservative    0;  Mismatches  156;  Indels    1;  Gaps    1;

Qy      19 CCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTC 78
           || | ||  | |      |           || || |   | || ||||    | | |
Db   83473 CCCCTCCCNCCTNNNCNNCNNNNNNNNNGNNCCCCCCCCCNNNCCCCCCCCCCCCCCCCCC 83532

Qy      79 CTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCC 138
           | |  || || || ||| | |  || || | |||  ||||  || | || || |   | |
Db   83533 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCNCCGC 83592

Qy     139 TAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTT 198
              |||  | | | |  || |  ||   |||| ||| || ||| | ||  ||||| |
Db   83593 CCCNCCCNCCCCCCCCNCCCCCGCCCGGGGCGCGGCCGCGGCCGCGGCGCGCGGCCCG 83652

Qy     199 GAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGC 258
           | | | | |||||||| |   |  | |  |||   |||   ||  ||| |  | |||||
Db   83653 CCGGGGGCGGGGGCGCCGGGGGGCGGGGCCGGGGCGGGGGGCG-GCGCGGGGGCGGCGGG 83711

Qy     259 CGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCA 318
           || || | | |  ||||| | ||| |     | | ||| ||||| |  |  ||   |||
Db   83712 CGCCCCGCGCGGCGCGGCGGGGGGCGGCGGCGGGGCGGGGCGGCGGNGGCCGGGCCCGCG 83771

Qy     319 ACGGGCGGCGGACGCTGCTGGGGAGGTGC 347
```

```
           |||  |   | ||  || || |  |  |
Db  83772 NCGGCGGCNCGGCGGGGCGGGCGCCGCNC 83800


RESULT 4
AC107145
LOCUS       AC107145               177168 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-106E18, *** SEQUENCING IN PROGRESS
            ***, 67 unordered pieces.
ACCESSION   AC107145
VERSION     AC107145.3  GI:21737159
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 177168)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177168)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 177168)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18701851.
```

```
---------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
    Center project name: GOQJ
    Center clone name: CH230-106E18
---------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 120304 bases at least Q40
    Consensus quality: 126448 bases at least Q30
    Consensus quality: 131535 bases at least Q20
----------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1020: contig of 1020 bp in length
*     1021     1120: gap of unknown length
*     1121     2478: contig of 1358 bp in length
*     2479     2578: gap of unknown length
*     2579     3589: contig of 1011 bp in length
*     3590     3689: gap of unknown length
*     3690     4826: contig of 1137 bp in length
*     4827     4926: gap of unknown length
*     4927     6143: contig of 1217 bp in length
*     6144     6243: gap of unknown length
*     6244     7578: contig of 1335 bp in length
*     7579     7678: gap of unknown length
*     7679     8692: contig of 1014 bp in length
*     8693     8792: gap of unknown length
*     8793     9896: contig of 1104 bp in length
*     9897     9996: gap of unknown length
*     9997    11336: contig of 1340 bp in length
*    11337    11436: gap of unknown length
*    11437    13217: contig of 1781 bp in length
*    13218    13317: gap of unknown length
*    13318    15162: contig of 1845 bp in length
*    15163    15262: gap of unknown length
*    15263    16473: contig of 1211 bp in length
*    16474    16573: gap of unknown length
*    16574    17582: contig of 1009 bp in length
*    17583    17682: gap of unknown length
*    17683    18900: contig of 1218 bp in length
*    18901    19000: gap of unknown length
*    19001    20097: contig of 1097 bp in length
*    20098    20197: gap of unknown length
*    20198    21776: contig of 1579 bp in length
*    21777    21876: gap of unknown length
*    21877    23292: contig of 1416 bp in length
*    23293    23392: gap of unknown length
*    23393    24939: contig of 1547 bp in length
*    24940    25039: gap of unknown length
*    25040    27311: contig of 2272 bp in length
*    27312    27411: gap of unknown length
*    27412    28968: contig of 1557 bp in length
*    28969    29068: gap of unknown length
*    29069    30742: contig of 1674 bp in length
*    30743    30842: gap of unknown length
*    30843    32409: contig of 1567 bp in length
*    32410    32509: gap of unknown length
*    32510    34377: contig of 1868 bp in length
*    34378    34477: gap of unknown length
*    34478    35582: contig of 1105 bp in length
*    35583    35682: gap of unknown length
*    35683    37613: contig of 1931 bp in length
*    37614    37713: gap of unknown length
*    37714    39152: contig of 1439 bp in length
*    39153    39252: gap of unknown length
*    39253    40589: contig of 1337 bp in length
*    40590    40689: gap of unknown length
*    40690    42099: contig of 1410 bp in length
*    42100    42199: gap of unknown length
*    42200    43365: contig of 1166 bp in length
*    43366    43465: gap of unknown length
*    43466    44469: contig of 1004 bp in length
*    44470    44569: gap of unknown length
*    44570    45636: contig of 1067 bp in length
*    45637    45736: gap of unknown length
*    45737    47826: contig of 2090 bp in length
*    47827    47926: gap of unknown length
*    47927    49153: contig of 1227 bp in length
*    49154    49253: gap of unknown length
*    49254    51098: contig of 1845 bp in length
*    51099    51198: gap of unknown length
*    51199    53492: contig of 2294 bp in length
*    53493    53592: gap of unknown length
*    53593    56557: contig of 2965 bp in length
*    56558    56657: gap of unknown length
*    56658    58011: contig of 1354 bp in length
*    58012    58111: gap of unknown length
*    58112    60168: contig of 2057 bp in length
*    60169    60268: gap of unknown length
*    60269    62494: contig of 2226 bp in length
*    62495    62594: gap of unknown length
*    62595    64041: contig of 1447 bp in length
*    64042    64141: gap of unknown length
*    64142    66152: contig of 2011 bp in length
*    66153    66252: gap of unknown length
*    66253    68551: contig of 2299 bp in length
*    68552    68651: gap of unknown length
*    68652    70360: contig of 1709 bp in length
*    70361    70460: gap of unknown length
*    70461    72984: contig of 2524 bp in length
*    72985    73084: gap of unknown length
*    73085    76398: contig of 3314 bp in length
*    76399    76498: gap of unknown length
*    76499    79360: contig of 2862 bp in length
*    79361    79460: gap of unknown length
*    79461    81701: contig of 2241 bp in length
*    81702    81801: gap of unknown length
*    81802    85349: contig of 3548 bp in length
*    85350    85449: gap of unknown length
*    85450    88601: contig of 3152 bp in length
*    88602    88701: gap of unknown length
*    88702    92201: contig of 3500 bp in length
*    92202    92301: gap of unknown length
*    92302    96407: contig of 4106 bp in length
*    96408    96507: gap of unknown length
*    96508   101699: contig of 5192 bp in length
*   101700   101799: gap of unknown length
*   101800   105571: contig of 3772 bp in length
*   105572   105671: gap of unknown length
*   105672   108750: contig of 3079 bp in length

  Query Match             15.9%;  Score 78;  DB 2;  Length 177168;
  Best Local Similarity   47.1%;  Pred. No. 7.6e-05;
  Matches  156;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy      19 CCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTC 78
           || |||           ||   | ||       | || ||  | |  ||||    | | |
Db   81318 CCCCCCNNNNCCCCNNCCCCNCCCCCNNNNCCNCCCCCCCCCCCCCNNCCCCCNCCCCCNC 81377

Qy      79 CTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCC 138
           | |  || || || ||| |    || || | |||  || |  || | || || |    ||
Db   81378 CCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCNCCNCCCCCCCCCCCCCCCCCGCC 81437
```

```
Qy     139 TAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTT 198
           |||     | || || ||     |  | || |  | |||  |     | ||||
Db   81438 CNNNNCCCNNNNCCCCCCCCCCCCNNNNCNCCCCCCCNNCCCCCNCGGGCCCGGCGGCGCN 81497

Qy     199 GAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGC 258
           |   ||| | |||| |   ||      |   ||   ||||| ||   | ||| ||||| ||
Db   81498 GGCGGCGCGNGGGCCCGGCCCCGGCGGGGGGGGGGGGGCCCCGNNGGGCGCGGGGGCCGC 81557

Qy     259 CGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCA 318
           |||| ||   |    ||  | |||||     ||| |   |  || || || |     |
Db   81558 CGGCGGGCNGGNGNGGGGCGGGGGGGGNCCCGCGCCCCGGGCGCGGGGGGGGCCCCGGGG 81617

Qy     319 ACGGGCGGCGGACGCTGCTGGGGAGGTGCTC 349
           || |    ||     |  ||| |  || |
Db   81618 CCGCGGCCGGGCNCNCGGGCGGGGGCNGCGC 81648
```

RESULT 5
AC111697
LOCUS AC111697 138467 bp DNA linear HTG 13-JUL-2002
DEFINITION Rattus norvegicus clone CH230-20C20, *** SEQUENCING IN PROGRESS ***, 59 unordered pieces.
ACCESSION AC111697
VERSION AC111697.2 GI:21736261
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciuromathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 138467)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Lear,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 138467)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 138467)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 12, 2002 this sequence version replaced gi:18701464.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- oject Information
Center project name: GMWR
Center clone name: CH230-20C20
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 97272 bases at least Q40
Consensus quality: 104545 bases at least Q30
Consensus quality: 110146 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1060: contig of 1060 bp in length
* 1061 1160: gap of unknown length
* 1161 2228: contig of 1068 bp in length
* 2229 2328: gap of unknown length
* 2329 3520: contig of 1192 bp in length
* 3521 3620: gap of unknown length
* 3621 4797: contig of 1177 bp in length
* 4798 4897: gap of unknown length
* 4898 6624: contig of 1727 bp in length
* 6625 6724: gap of unknown length
* 6725 8301: contig of 1577 bp in length
* 8302 8401: gap of unknown length
* 8402 9813: contig of 1412 bp in length
* 9814 9913: gap of unknown length
* 9914 11128: contig of 1215 bp in length
* 11129 11228: gap of unknown length
* 11229 12658: contig of 1430 bp in length
* 12659 12758: gap of unknown length
* 12759 15360: contig of 2602 bp in length
* 15361 15460: gap of unknown length
* 15461 17025: contig of 1565 bp in length
* 17026 17125: gap of unknown length
* 17126 18259: contig of 1134 bp in length
* 18260 18359: gap of unknown length
* 18360 19979: contig of 1620 bp in length
* 19980 20079: gap of unknown length
* 20080 21208: contig of 1129 bp in length
* 21209 21308: gap of unknown length
* 21309 22587: contig of 1279 bp in length
* 22588 22687: gap of unknown length
* 22688 24387: contig of 1700 bp in length
* 24388 24487: gap of unknown length
* 24488 26131: contig of 1644 bp in length
* 26132 26231: gap of unknown length

```
*    26232    27699: contig of 1468 bp in length
*    27700    27799: gap of unknown length
*    27800    29145: contig of 1346 bp in length
*    29146    29245: gap of unknown length
*    29246    30337: contig of 1092 bp in length
*    30338    30437: gap of unknown length
*    30438    31510: contig of 1073 bp in length
*    31511    31610: gap of unknown length
*    31611    33147: contig of 1537 bp in length
*    33148    33247: gap of unknown length
*    33248    34743: contig of 1496 bp in length
*    34744    34843: gap of unknown length
*    34844    36324: contig of 1481 bp in length
*    36325    36424: gap of unknown length
*    36425    37988: contig of 1564 bp in length
*    37989    38088: gap of unknown length
*    38089    40522: contig of 2434 bp in length
*    40523    40622: gap of unknown length
*    40623    41876: contig of 1254 bp in length
*    41877    41976: gap of unknown length
*    41977    43465: contig of 1489 bp in length
*    43466    43565: gap of unknown length
*    43566    45265: contig of 1700 bp in length
*    45266    45365: gap of unknown length
*    45366    47618: contig of 2253 bp in length
*    47619    47718: gap of unknown length
*    47719    49624: contig of 1906 bp in length
*    49625    49724: gap of unknown length
*    49725    51701: contig of 1977 bp in length
*    51702    51801: gap of unknown length
*    51802    54174: contig of 2373 bp in length
*    54175    54274: gap of unknown length
*    54275    56347: contig of 2073 bp in length
*    56348    56447: gap of unknown length
*    56448    57737: contig of 1290 bp in length
*    57738    57837: gap of unknown length
*    57838    60324: contig of 2487 bp in length
*    60325    60424: gap of unknown length
*    60425    61731: contig of 1307 bp in length
*    61732    61831: gap of unknown length
*    61832    63545: contig of 1714 bp in length
*    63546    63645: gap of unknown length
*    63646    65262: contig of 1617 bp in length
*    65263    65362: gap of unknown length
*    65363    67811: contig of 2449 bp in length
*    67812    67911: gap of unknown length
*    67912    70170: contig of 2259 bp in length
*    70171    70270: gap of unknown length
*    70271    72210: contig of 1940 bp in length
*    72211    72310: gap of unknown length
*    72311    73697: contig of 1387 bp in length
*    73698    73797: gap of unknown length
*    73798    75398: contig of 1601 bp in length
*    75399    75498: gap of unknown length
*    75499    78702: contig of 3204 bp in length
*    78703    78802: gap of unknown length
*    78803    81340: contig of 2538 bp in length
*    81341    81440: gap of unknown length
*    81441    84907: contig of 3467 bp in length
*    84908    85007: gap of unknown length
*    85008    88196: contig of 3189 bp in length
*    88197    88296: gap of unknown length
*    88297    91902: contig of 3606 bp in length
*    91903    92002: gap of unknown length
*    92003    95117: contig of 3115 bp in length
*    95118    95217: gap of unknown length
*    95218    98243: contig of 3026 bp in length
*    98244    98343: gap of unknown length
*    98344   102722: contig of 4379 bp in length
*   102723   102822: gap of unknown length
*   102823   105855: contig of 3033 bp in length
*   105856   105955: gap of unknown length
*   105956   110171: contig of 4216 bp in length
```

```
  Query Match             15.6%;  Score 76.8;  DB 2;  Length 138467;
  Best Local Similarity   52.4%;  Pred. No. 0.00013;
  Matches  176;  Conservative    0;  Mismatches  159;  Indels    1;  Gaps    1;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | || ||||           ||     |||  |  || || ||  | || ||||    | |
Db    2848 CNCCCCCCCCCNCCCCCNCCCNNNNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 2907

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
             |||  || || || ||| | |  || || | |||  ||||  |  | || || |   |
Db    2908 CTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCC 2967

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | | || || ||  | || | |||||    ||| |     || | |
Db    2968 CCCCGCCCCCCCCCCCCCCCCCCCCCGCCCCNCCCCGCCCNCGCCGGGGGCGGCGGGCGC 3027

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
             ||||   ||||| | |  |  | |   ||   | | | ||  |||||     ||
Db    3028 CGCGCCGCGCCGGGGCGGCCGGGGCGGGGCGGGGCCGGCGGCCCGCCCGCCGGCCCCGCC 3087

Qy     257 GCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCG 316
           | || |    | || || || | ||||| |  || | | |  | || || || || || | |
Db    3088 GGCGCCGCCCGGGCCGC-GCGGGGGGGGCGGCGGGGGCCGCCCCGCGGGCGGGGGGGGGG 3146

Qy     317 CAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANG 352
             ||||||||  |  |    || || || |  |
Db    3147 GGCGCGGCGGCGGCGGGGGGGCCGGGGGGGCGCGCG 3182
```

```
RESULT 6
AC062001
LOCUS       AC062001               82815 bp    DNA     linear   HTG 21-APR-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-124H2 map 4, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC062001
VERSION     AC062001.1  GI:7630740
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 82815)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-124H2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82815)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
```

```
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L9871
    Center clone name: 124_H_2
--------------.
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      975: contig of 975 bp in length
*      976 1075: gap of      100 bp
*     1076     2100: contig of 1025 bp in length
*     2101 2200: gap of      100 bp
*     2201     3185: contig of 985 bp in length
*     3186 3285: gap of      100 bp
*     3286     4151: contig of 866 bp in length
*     4152 4251: gap of      100 bp
*     4252     5206: contig of 955 bp in length
*     5207 5306: gap of      100 bp
*     5307     6287: contig of 981 bp in length
*     6288 6387: gap of      100 bp
*     6388     7332: contig of 945 bp in length
*     7333 7432: gap of      100 bp
*     7433     8397: contig of 965 bp in length
*     8398 8497: gap of      100 bp
*     8498     9457: contig of 960 bp in length
*     9458 9557: gap of      100 bp
*     9558    10489: contig of 932 bp in length
*    10490 10589: gap of      100 bp
*    10590    11602: contig of 1013 bp in length
*    11603 11702: gap of      100 bp
*    11703    12694: contig of 992 bp in length
*    12695 12794: gap of      100 bp
*    12795    13798: contig of 1004 bp in length
*    13799 13898: gap of      100 bp
*    13899    14866: contig of 968 bp in length
*    14867 14966: gap of      100 bp
*    14967    15930: contig of 964 bp in length
*    15931 16030: gap of      100 bp
*    16031    16981: contig of 951 bp in length
*    16982 17081: gap of      100 bp
*    17082    18043: contig of 962 bp in length
*    18044 18143: gap of      100 bp
*    18144    19119: contig of 976 bp in length
*    19120 19219: gap of      100 bp
*    19220    20173: contig of 954 bp in length
*    20174 20273: gap of      100 bp
*    20274    21232: contig of 959 bp in length
*    21233 21332: gap of      100 bp
*    21333    22227: contig of 895 bp in len th
*    22228 22327: gap of      100 bp
*    22328    23303: contig of 976 bp in length
*    23304 23403: gap of      100 bp
*    23404    24393: contig of 990 bp in length
*    24394 24493: gap of      100 bp
*    24494    25495: contig of 1002 bp in length
*    25496 25595: gap of      100 bp
*    25596    26554: contig of 959 bp in length
*    26555 26654: gap of      100 bp
*    26655    27611: contig of 957 bp in length
*    27612 27711: gap of      100 bp
*    27712    28689: contig of 978 bp in length
*    28690 28789: gap of      100 bp
*    28790    29750: contig of 961 bp in length
*    29751 29850: gap of      100 bp
*    29851    30790: contig of 940 bp in length
*    30791 30890: gap of      100 bp
*    30891    31881: contig of 991 bp in length
*    31882 31981: gap of      100 bp
*    31982    32962: contig of 981 bp in length
*    32963 33062: gap of      100 bp
*    33063    34046: contig of 984 bp in length
*    34047 34146: gap of      100 bp
*    34147    35108: contig of 962 bp in length
*    35109 35208: gap of      100 bp
*    35209    36136: contig of 928 bp in length
*    36137 36236: gap of      100 bp
*    36237    37202: contig of 966 bp in length
*    37203 37302: gap of      100 bp
*    37303    38272: contig of 970 bp in length
*    38273 38372: gap of      100 bp
*    38373    39319: contig of 947 bp in length
*    39320 39419: gap of      100 bp
*    39420    40376: contig of 957 bp in length
*    40377 40476: gap of      100 bp
*    40477    41464: contig of 988 bp in length
*    41465 41564: gap of      100 bp
*    41565    42492: contig of 928 bp in length
*    42493 42592: gap of      100 bp
*    42593    43578: contig of 986 bp in length
*    43579 43678: gap of      100 bp
*    43679    44686: contig of 1008 bp in length
*    44687 44786: gap of      100 bp
*    44787    45745: contig of 959 bp in length
*    45746 45845: gap of      100 bp
*    45846    46827: contig of 982 bp in length
*    46828 46927: gap of      100 bp
*    46928    47910: contig of 983 bp in length
*    47911 48010: gap of      100 bp
*    48011    48969: contig of 959 bp in length
*    48970 49069: gap of      100 bp
*    49070    50028: contig of 959 bp in length
*    50029 50128: gap of      100 bp
*    50129    51053: contig of 925 bp in length
*    51054 51153: gap of      100 bp
*    51154    52111: contig of 958 bp in length
*    52112 52211: gap of      100 bp
*    52212    53189: contig of 978 bp in length
*    53190 53289: gap of      100 bp
*    53290    54271: contig of 982 bp in length
*    54272 54371: gap of      100 bp
*    54372    55336: contig of 965 bp in length
*    55337 55436: gap of      100 bp
*    55437    56366: contig of 930 bp in length
*    56367 56466: gap of      100 bp
*    56467    57415: contig of 949 bp in length
*    57416 57515: gap of      100 bp
*    57516    58501: contig of 986 bp in length
*    58502 58601: gap of      100 bp
*    58602    59567: contig of 966 bp in length
*    59568 59667: gap of      100 bp
*    59668    60585: contig of 918 bp in length
*    60586 60685: gap of      100 bp
*    60686    61646: contig of 961 bp in length
*    61647 61746: gap of      100 bp
*    61747    62739: contig of 993 bp in length
*    62740 62839: gap of      100 bp
*    62840    63824: contig of 985 bp in length
*    63825 63924: gap of      100 bp
*    63925    64903: contig of 979 bp in length
*    64904 65003: gap of      100 bp
*    65004    65963: contig of 960 bp in length
*    65964 66063: gap of      100 bp
*    66064    67010: contig of 947 bp in length
```

```
            *    67011 67110: gap of      100 bp
            *    67111    68077: contig of 967 bp in length
            *    68078 68177: gap of      100 bp
            *    68178    69147: contig of 970 bp in length
            *    69148 69247: gap of      100 bp
            *    69248    70258: contig of 1011 bp in length
            *    70259 70358: gap of      100 bp
            *    70359    71286: contig of 928 bp in length
            *    71287 71386: gap of      100 bp
            *    71387    72279: contig of 893 bp in length
            *    72280 72379: gap of      100 bp
            *    72380    73317: contig of 938 bp in length
            *    73318 73417: gap of      100 bp

  Query Match             15.1%;  Score 74.4;  DB 2;  Length 82815;
  Best Local Similarity   48.9%;  Pred. No. 0.00036;
  Matches  159;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | || |||| |         ||   | | |  |  || || ||  | || | ||    | |
Db   66610 CCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNCCCCCCCCCCCCCCCCNCCNCCCCCC 66669

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
            || |  || || || ||| | |   | || | |||  |||| || | |  || |
Db   66670 CCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCGCN 66729

Qy     137 CCTAACCCCGCTCTCGCCTCCTCCGCCTCCGCGCCCCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | | || || || ||    | ||    | || |    |   ||||
Db   66730 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNNCCCCCNNNCCCCGGCCCCNNNCGCGNNN 66789

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
           | | ||| | || |     | |    |   |    || |  ||  |||| |    |||
Db   66790 CGGCGGGCGCGGGGCCCGGGCCGGGCNGGGGNGCGCCGGCCGCCGGGCGCCCCGCCCGCG 66849

Qy     257 GCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCG 316
           |  ||| || | ||    ||   |||||     ||  |||  |||  || |  || || |
Db   66850 GGGGGCGGGGGGGCGCGCGCCCCGGGGGNNGCCGCCGCCGCTCGGGCGGGGCGGGGGGGG 66909

Qy     317 CAACGGGCGGCGGACGCTGCTGGGG 341
           |     |||  |  |   || || |
Db   66910 CCGGCNGCGCGGCCCCGCGCCGGCG 66934


RESULT 7
AC121439
LOCUS       AC121439               158897 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-247E4, *** SEQUENCING IN PROGRESS
            ***, 56 unordered pieces.
ACCESSION   AC121439
VERSION     AC121439.2  GI:21321741
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 158897)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158897)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 158897)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 4, 2002 this sequence version replaced gi:20976389.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: GXXP
                Center clone name: CH230-247E4
            ------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 125365 bases at least Q40
                Consensus quality: 129911 bases at least Q30
                Consensus quality: 132859 bases at least Q20
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 56 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1137: contig of 1137 bp in length
            *     1138     1237: gap of unknown length
            *     1238     2390: contig of 1153 bp in length
            *     2391     2490: gap of unknown length
            *     2491     3583: contig of 1093 bp in length
            *     3584     3683: gap of unknown length
            *     3684     4820: contig of 1137 bp in length
            *     4821     4920: gap of unknown length
```

```
*     4921     6335: contig of 1415 bp in length
*     6336     6435: gap of unknown length
*     6436     7534: contig of 1099 bp in length
*     7535     7634: gap of unknown length
*     7635     9844: contig of 2210 bp in length
*     9845     9944: gap of unknown length
*     9945    11234: contig of 1290 bp in length
*    11235    11334: gap of unknown length
*    11335    12344: contig of 1010 bp in length
*    12345    12444: gap of unknown length
*    12445    13678: contig of 1234 bp in length
*    13679    13778: gap of unknown length
*    13779    15477: contig of 1699 bp in length
*    15478    15577: gap of unknown length
*    15578    16737: contig of 1160 bp in length
*    16738    16837: gap of unknown length
*    16838    19340: contig of 2503 bp in length
*    19341    19440: gap of unknown length
*    19441    20568: contig of 1128 bp in length
*    20569    20668: gap of unknown length
*    20669    22352: contig of 1684 bp in length
*    22353    22452: gap of unknown length
*    22453    23795: contig of 1343 bp in length
*    23796    23895: gap of unknown length
*    23896    25953: contig of 2058 bp in length
*    25954    26053: gap of unknown length
*    26054    28136: contig of 2083 bp in length
*    28137    28236: gap of unknown length
*    28237    30151: contig of 1915 bp in length
*    30152    30251: gap of unknown length
*    30252    31447: contig of 1196 bp in length
*    31448    31547: gap of unknown length
*    31548    32949: contig of 1402 bp in length
*    32950    33049: gap of unknown length
*    33050    34438: contig of 1389 bp in length
*    34439    34538: gap of unknown length
*    34539    36281: contig of 1743 bp in length
*    36282    36381: gap of unknown length
*    36382    37951: contig of 1570 bp in length
*    37952    38051: gap of unknown length
*    38052    39294: contig of 1243 bp in length
*    39295    39394: gap of unknown length
*    39395    40924: contig of 1530 bp in length
*    40925    41024: gap of unknown length
*    41025    42151: contig of 1127 bp in length
*    42152    42251: gap of unknown length
*    42252    43924: contig of 1673 bp in length
*    43925    44024: gap of unknown length
*    44025    46633: contig of 2609 bp in length
*    46634    46733: gap of unknown length
*    46734    48320: contig of 1587 bp in length
*    48321    48420: gap of unknown length
*    48421    50491: contig of 2071 bp in length
*    50492    50591: gap of unknown length
*    50592    53057: contig of 2466 bp in length
*    53058    53157: gap of unknown length
*    53158    56233: contig of 3076 bp in length
*    56234    56333: gap of unknown length
*    56334    59001: contig of 2668 bp in length
*    59002    59101: gap of unknown length
*    59102    61813: contig of 2712 bp in length
*    61814    61913: gap of unknown length
*    61914    66256: contig of 4343 bp in length
*    66257    66356: gap of unknown length
*    66357    68713: contig of 2357 bp in length
*    68714    68813: gap of unknown length
*    68814    70772: contig of 1959 bp in length
*    70773    70872: gap of unknown length
*    70873    74228: contig of 3356 bp in length
*    74229    74328: gap of unknown length
*    74329    77316: contig of 2988 bp in length
*    77317    77416: gap of unknown length
*    77417    81174: contig of 3758 bp in length
*    81175    81274: gap of unknown length
*    81275    85927: contig of 4653 bp in length
*    85928    86027: gap of unknown length
*    86028    89056: contig of 3029 bp in length
*    89057    89156: gap of unknown length
*    89157    93557: contig of 4401 bp in length
*    93558    93657: gap of unknown length
*    93658    96490: contig of 2833 bp in length
*    96491    96590: gap of unknown length
*    96591   101020: contig of 4430 bp in length
*   101021   101120: gap of unknown length
*   101121   104653: contig of 3533 bp in length
*   104654   104753: gap of unknown length
*   104754   108393: contig of 3640 bp in length
*   108394   108493: gap of unknown length
*   108494   112698: contig of 4205 bp in length
*   112699   112798: gap of unknown length
*   112799   118578: contig of 5780 bp in length
*   118579   118678: gap of unknown length
*   118679   124462: contig of 5784 bp in length
*   124463   124562: gap of unknown length
*   124563   129728: contig of 5166 bp in length
*   129729   129828: gap of unknown length
*   129829   135194: contig of 5366 bp in length
*   135195   135294: gap of unknown length
*   135295   143870: contig of 8576 bp in length

  Query Match             14.9%;  Score 73.2;  DB 2;  Length 158897;
  Best Local Similarity   49.9%;  Pred. No. 0.00058;
  Matches  183;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy      10 CGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCC 69
           | || | |||  | ||  | ||    |||    | |  |  || || ||  ||| |||||
Db   35832 CTTCTTCCTCTCCTCCTTCCTTCCTTCCTCCTCCTCTCCCTCCCCCCCCCCCCTCTTCCCC 35891

Qy      70 AATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTC 129
             ||  | || ||||||||| ||| |||   | | || | || |||   | |  ||||
Db   35892 CCCTCTTCCTTCCTCCTCCTCCCTCTCTCTCCCTTCTCTCTCCCCTTTCTTTCCCTC 35951

Qy     130 CGCAAACCCTAACCCC( CTCTCGCCTCCTCCGCCTCCGCG CGCCTCGCCCGTGTCGTC 189
             |    |||   |||; ||||| || | || || |  |  | ||||  |  | |  |
Db   35952 TTCTTCTC :TCCTCCCCTCTCTCCTCTGCCCCTCCCCTCCTTCCCCTCCTCTTTCTTCCC 36011

Qy     190 CGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGC 249
             || |  ||     |   |   |          | |    |   |||  | || |
Db   36012 TTCGCTGCTGGCGC:CCCTTCTCTGTCCGGTCGGCGGCGCGGCGGGGGCGCGGTGTTGTG 36071

Qy     250 TGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGA 309
           | |  |  || |||  | |  | ||  | ||| | |  || |  || |  || | ||| ||||
Db   36072 GGTGCGGCTCGTCCGCTGCGGCGGTGA' 'AGGGCGTGCGGGGGGGGGGGGGGCGGGTGTG 36131

Qy     310 GGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCG 369
           || || |    |||  | ||     | ||||| |  | ||  |   |      | | |||
Db   36132 GGGGGGGGTTGGGGGTGAGGGGCGGGGTGGGGGGTGGGTCGCGCGCTGTGGTGGGCGGCG 36191

Qy     370 GGTTGAG 376
           ||| | |
Db   36192 GGTGGCG 36198


RESULT 8
AC116109
LOCUS       AC116109               68330 bp    DNA     linear   HTG 25-MAR-2002
DEFINITION  Mus musculus clone RP23-467G14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC116109
VERSION     AC116109.1  GI:19703220
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 68330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
  TITLE     Mus musculus, clone RP23-467G14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22810
                Center clone name: 467_G_14
            --------------.
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      647: contig of 647 bp in length
*      648 747: gap of      100 bp
*      748     1422: contig of 675 bp in length
*     1423 1522: gap of      100 bp
*     1523     2229: contig of 707 bp in length
*     2230 2329: gap of      100 bp
*     2330     3001: contig of 672 bp in length
*     3002 3101: gap of      100 bp
*     3102     3783: contig of 682 bp in length
*     3784 3883: gap of      100 bp
*     3884     4586: contig of 703 bp in length
*     4587 4686: gap of      100 bp
*     4687     5381: contig of 695 bp in length
*     5382 5481: gap of      100 bp
*     5482     6190: contig of 709 bp in length
*     6191 6290: gap of      100 bp
*     6291     7003: contig of 713 bp in length
*     7004 7103: gap of      100 bp
*     7104     7824: contig of 721 bp in length
*     7825 7924: gap of      100 bp
*     7925     8662: contig of 738 bp in length
*     8663 8762: gap of      100 bp
*     8763     9402: contig of 640 bp in length
*     9403 9502: gap of      100 bp
*     9503    10235: contig of 733 bp in length
*    10236 10335: gap of      100 bp
*    10336    11055: contig of 720 bp in length
*    11056 11155: gap of      100 bp
*    11156    11861: contig of 706 bp in length
*    11862 11961: gap of      100 bp
*    11962    12653: contig of 692 bp in length
*    12654 12753: gap of      100 bp
*    12754    13445: contig of 692 bp in length
*    13446 13545: gap of      100 bp
*    13546    14235: contig of 690 bp in length
*    14236 14335: gap of      100 bp
*    14336    15046: contig of 711 bp in length
*    15047 15146: gap of      100 bp
*    15147    15853: contig of 707 bp in length
*    15854 159[illegible]: gap of      100 bp
*    15954    [illegible]632: contig of 679 bp in length
*    16633 16732: gap of      100 bp
*    16733    17441: contig of 709 bp in length
*    17442 17541: gap of      100 bp
*    17542    18270: contig of 729 bp in length
*    18271 18370: gap of      100 bp
*    18371    19048: contig of 678 bp in length
*    19049 19148: gap of      100 bp
*    19149    19851: contig of 703 bp in length
*    19852 19951: gap of      100 bp
*    19952    20661: contig of 710 bp in length
*    20662 20761: gap of      100 bp
*    20762    21473: contig of 712 bp in length
*    21474 21573: gap of      100 bp
*    21574    22286: contig of 713 bp in length
*    22287 22386: gap of      100 bp
*    22387    23086: contig of 700 bp in length
*    23087 23186: gap of      100 bp
*    23187    23920: contig of 734 bp in length
*    23921 24020: gap of      100 bp
*    24021    24713: contig of 693 bp in length
*    24714 24813: gap of      100 bp
*    24814    25540: contig of 727 bp in length
*    25541 25640: gap of      100 bp
*    25641    26364: contig of 724 bp in length
*    26365 26464: gap of      100 bp
*    26465    27180: contig of 716 bp in length
*    27181 27280: gap of      100 bp
*    27281    27977: contig of 697 bp in length
*    27978 28077: gap of      100 bp
*    28078    28769: contig of 692 bp in length
*    28770 28869: gap of      100 bp
*    28870    29559: contig of 690 bp in length
*    29560 29659: gap of      100 bp
*    29660    30357: contig of 698 bp in length
*    30358 30457: gap of      100 bp
*    30458    31165: contig of 708 bp in length
*    31166 31265: gap of      100 bp
*    31266    31948: contig of 683 bp in length
*    31949 32048: gap of      100 bp
*    32049    32768: contig of 720 bp in length
*    32769 32868: gap of      100 bp
*    32869    33548: contig of 680 bp in length
*    33549 33648: gap of      100 bp
*    33649    34357: contig of 709 bp in length
*    34358 34457: gap of      100 bp
*    34458    35187: contig of 730 bp in length
*    35188 35287: gap of      100 bp
*    35288    35990: contig of 703 bp in length
*    35991 36090: gap of      100 bp
*    36091    36767: contig of 677 bp in length
*    36768 36867: gap of      100 bp
*    36868    37575: contig of 708 bp in length
*    37576 37675: gap of      100 bp
*    37676    38379: contig of 704 bp in length
```

```
*    38380 38479: gap of      100 bp
*    38480    39200: contig of 721 bp in length
*    39201 39300: gap of      100 bp
*    39301    40023: contig of 723 bp in length
*    40024 40123: gap of      100 bp
*    40124    40827: contig of 704 bp in length
*    40828 40927: gap of      100 bp
*    40928    41659: contig of 732 bp in length
*    41660 41759: gap of      100 bp
*    41760    42482: contig of 723 bp in length
*    42483 42582: gap of      100 bp
*    42583    43307: contig of 725 bp in length
*    43308 43407: gap of      100 bp
*    43408    44124: contig of 717 bp in length
*    44125 44224: gap of      100 bp
*    44225    44923: contig of 699 bp in length
*    44924 45023: gap of      100 bp
*    45024    45713: contig of 690 bp in length
*    45714 45813: gap of      100 bp
*    45814    46506: contig of 693 bp in length
*    46507 46606: gap of      100 bp
*    46607    47291: contig of 685 bp in length
*    47292 47391: gap of      100 bp
*    47392    48100: contig of 709 bp in length
*    48101 48200: gap of      100 bp
*    48201    48910: contig of 710 bp in length
*    48911 49010: gap of      100 bp
*    49011    49727: contig of 717 bp in length
*    49728 49827: gap of      100 bp
*    49828    50558: contig of 731 bp in length
*    50559 50658: gap of      100 bp
*    50659    51335: contig of 677 bp in length
*    51336 51435: gap of      100 bp
*    51436    52168: contig of 733 bp in length
*    52169 52268: gap of      100 bp
*    52269    52992: contig of 724 bp in length
*    52993 53092: gap of      100 bp
*    53093    53810: contig of 718 bp in length
*    53811 53910: gap of      100 bp
*    53911    54559: contig of 649 bp in length
*    54560 54659: gap of      100 bp
*    54660    55361: contig of 702 bp in length
*    55362 55461: gap of      100 bp

  Query Match             14.7%;  Score 72.2;  DB 2;  Length 68330;
  Best Local Similarity   52.3%;  Pred. No. 0.00093;
  Matches  137;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy       35 ACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACC 94
           ||       |     |  ||  |  | || ||||    | | || |  || || || ||
Db      304 ACNNNNNNNCNNCNCNCCCCNCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCC 363

Qy       95 CACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGC 154
           | | |  || || | |||  ||||  || | || || |   |||   ||||| | | | |
Db      364 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCC 423

Qy      155 CTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGC 214
           | || || || || | || || | | |     | | | ||    | |||||    |||||
Db      424 CCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCGGGGCGGGGGCGCGGGGCGCGCCCCGGCGC 483

Qy      215 CAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCG 274
              |||  | |   ||   |||    |   |  || || |||||||| ||| | |  |
Db      484 GGCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGG:GCGGCCGGGCGGCGGGGGGGC 543

Qy      275 GCAGTGGGGGAGTTGGCGACGG 296
           |  |  |||| |    |  |||
Db      544 GGNGGNGGGGGGCCCCCCGCGG 565


RESULT 9
AC126076
LOCUS       AC126076               117082 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-143P6, *** SEQUENCING IN PROGRESS
            ***, 63 unordered pieces.
ACCESSION   AC126076
VERSION     AC126076.2  GI:21699142
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 117082)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 117082)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 117082)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 5, 2002 this sequence version replaced gi:21686820.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GZTG
                Center clone name: CH230-143P6
```

```
-------------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 55139 bases at least Q40
    Consensus quality: 61060 bases at least Q30
    Consensus quality: 64630 bases at least Q20
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1055: contig of 1055 bp in length
*     1056     1155: gap of unknown length
*     1156     2165: contig of 1010 bp in length
*     2166     2265: gap of unknown length
*     2266     3314: contig of 1049 bp in length
*     3315     3414: gap of unknown length
*     3415     5018: contig of 1604 bp in length
*     5019     5118: gap of unknown length
*     5119     6243: contig of 1125 bp in length
*     6244     6343: gap of unknown length
*     6344     7662: contig of 1319 bp in length
*     7663     7762: gap of unknown length
*     7763     9164: contig of 1402 bp in length
*     9165     9264: gap of unknown length
*     9265    10352: contig of 1088 bp in length
*    10353    10452: gap of unknown length
*    10453    11637: contig of 1185 bp in length
*    11638    11737: gap of unknown length
*    11738    13085: contig of 1348 bp in length
*    13086    13185: gap of unknown length
*    13186    14617: contig of 1432 bp in length
*    14618    14717: gap of unknown length
*    14718    16407: contig of 1690 bp in length
*    16408    16507: gap of unknown length
*    16508    17676: contig of 1169 bp in length
*    17677    17776: gap of unknown length
*    17777    19207: contig of 1431 bp in length
*    19208    19307: gap of unknown length
*    19308    21017: contig of 1710 bp in length
*    21018    21117: gap of unknown length
*    21118    22172: contig of 1055 bp in length
*    22173    22272: gap of unknown length
*    22273    23366: contig of 1094 bp in length
*    23367    23466: gap of unknown length
*    23467    24549: contig of 1083 bp in length
*    24550    24649: gap of unknown length
*    24650    25665: contig of 1016 bp in length
*    25666    25765: gap of unknown length
*    25766    27421: contig of 1656 bp in length
*    27422    27521: gap of unknown length
*    27522    28564: contig of 1043 bp in length
*    28565    28664: gap of unknown length
*    28665    29755: contig of 1091 bp in length
*    29756    29855: gap of unknown length
*    29856    31657: contig of 1802 bp in length
*    31658    31757: gap of unknown length
*    31758    33595: contig of 1838 bp in length
*    33596    33695: gap of unknown length
*    33696    35198: contig of 1503 bp in length
*    35199    35298: gap of unknown length
*    35299    36466: contig of 1168 bp in length
*    36467    36566: gap of unknown length
*    36567    37640: contig of 1074 bp in length
*    37641    37740: gap of unknown length
*    37741    39003: contig of 1263 bp in length
*    39004    39103: gap of unknown length
*    39104    41305: contig of 2202 bp in length
*    41306    41405: gap of unknown length
*    41406    42620: contig of 1215 bp in length
*    42621    42720: gap of unknown length
*    42721    43940: contig of 1220 bp in length
*    43941    44040: gap of unknown length
*    44041    45996: contig of 1956 bp in length
*    45997    46096: gap of unknown length
*    46097    47529: contig of 1433 bp in length
*    47530    47629: gap of unknown length
*    47630    49451: contig of 1822 bp in length
*    49452    49551: gap of unknown length
*    49552    50801: contig of 1250 bp in length
*    50802    50901: gap of unknown length
*    50902    52881: contig of 1980 bp in length
*    52882    52981: gap of unknown length
*    52982    54499: contig of 1518 bp in length
*    54500    54599: gap of unknown length
*    54600    55934: contig of 1335 bp in length
*    55935    56034: gap of unknown length
*    56035    58138: contig of 2104 bp in length
*    58139    58238: gap of unknown length
*    58239    59732: contig of 1494 bp in length
*    59733    59832: gap of unknown length
*    59833    61285: contig of 1453 bp in length
*    61286    61385: gap of unknown length
*    61386    62880: contig of 1495 bp in length
*    62881    62980: gap of unknown length
*    62981    64930: contig of 1950 bp in length
*    64931    65030: gap of unknown length
*    65031    66998: contig of 1968 bp in length
*    66999    67098: gap of unknown length
*    67099    69768: contig of 2670 bp in length
*    69769    69868: gap of unknown length
*    69869    72122: contig of 2254 bp in length
*    72123    72222: gap of unknown length
*    72223    74144: contig of 1922 bp in length
*    74145    74244: gap of unknown length
*    74245    75957: contig of 1713 bp in length
*    75958    76057: gap of unknown length
*    76058    78577: contig of 2520 bp in length
*    78578    78677: gap of unknown length
*    78678    80255: contig of 1578 bp in length
*    80256    80355: gap of unknown length
*    80356    82593: contig of 2238 bp in length
*    82594    82693: gap of unknown length
*    82694    84685: contig of 1992 bp in length
*    84686    84785: gap of unknown length
*    84786    86448: contig of 1663 bp in length
*    86449    86548: gap of unknown length
*    86549    87884: contig of 1336 bp in length

  Query Match          14.7%;  Score 72.2;  DB 2;  Length 117082;
  Best Local Similarity  44.6%;  Pred. No. 0.00091;
  Matches  152;  Conservative    0;  Mismatches  189;  Indels    0;  Gaps    0;

Qy       9 GCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCC 68
           |                                                  ||  | ||  | || |||
Db   48749 GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCNCCCCCCCCCCCCCCC 48808

Qy      69 CAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCT 128
           |       | ||    || || |  ||| | |  || ||    ||  |||   || | ||
Db   48809 CNNCAAACCCCCNNNCCCCCCCNCCCCCCCCCCCCCCCCCCNNNCCGCCCCGCCCCCCCCCC 48868

Qy     129 CCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGT 188
           || |   |||   | ||| | | | ||  | | ||| ||||| || |      | | ||
Db   48869 CCCCCGCCCCCCCCCNCCCCCCCCCCCCCNNCCCGGCCGCCGCGGCGGCGNCGNGGCGGCGG 48928

Qy     189 CCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCG 248
           | |  | |  | ||| | | ||| |   |  | | | |   ||   ||  | |   | |
Db   48929 CGGGCGCCGCGCGCGGGCGGGGGGGGGGGGGGGCGGGGGGGGGGGGCGCGGGGGGGGG 48988
```

```
Qy     249 CTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGG 308
           |||| ||  || ||| | |   |||    || || |  ||||  ||   ||     ||
Db   48989 GCGGGGGGGGGGGGCGGGGGGGGCCGGGGCGGGCGGGGGGGGCGGGGGGCGGGGCCCGGG 49048

Qy     309 AGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTC 349
            ||  | |    ||| || ||  || || |||   | ||||
Db   49049 CGGGCGGGGGGGGGGGGGGGGGGCGGCGGGGCGCGCGCTC 49089


RESULT 10
AC027353/c
LOCUS       AC027353               101509 bp    DNA     linear   HTG 26-SEP-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-167D21, WORKING DRAFT
            SEQUENCE, 68 unordered pieces.
ACCESSION   AC027353
VERSION     AC027353.4  GI:10305180
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 101509)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101509)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi 9090641.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-11_167D21
            -------
            Summary Statistics
            Consensus quality: 19476 bases at least Q40
            Consensus quality: 22258 bases at least Q30
            Consensus quality: 25343 bases at least Q20
            Estimated insert size: 26000; agarose-fp estimation
            Estimated insert size: 94809; sum-of-contigs estimation
            Quality coverage:  15.03 in Q20 bases; agarose-fp estimation
            Quality coverage:  4.12 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 68 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1470: contig of 1470 bp in length
            *     1471     1570: gap of unknown length
            *     1571     2728: contig of 1158 bp in length
            *     2729     2828: gap of unknown length
            *     2829     3955: contig of 1127 bp in length
            *     3956     4055: gap of unknown length
            *     4056     5273: contig of 1218 bp in length
            *     5274     5373: gap of unknown length
            *     5374     6455: contig of 1082 bp in length
            *     6456     6555: gap of unknown length
            *     6556     7722: contig of 1167 bp in length
            *     7723     7822: gap of unknown length
            *     7823     8994: contig of 1172 bp in length
            *     8995     9094: gap of unknown length
            *     9095    10258: contig of 1164 bp in length
            *    10259    10358: gap of unknown length
            *    10359    11529: contig of 1171 bp in length
            *    11530    11629: gap of unknown length
            *    11630    12772: contig of 1143 bp in length
            *    12773    12872: gap of unknown length
            *    12873    13954: contig of 1082 bp in length
            *    13955    14054: gap of unknown length
            *    14055    15127: contig of 1073 bp in length
            *    15128    15227: gap of unknown length
            *    15228    16411: contig of 1184 bp in length
            *    16412    16511: gap of unknown length
            *    16512    17557: contig of 1046 bp in length
            *    17558    17657: gap of unknown length
            *    17658    18766: contig of 1109 bp in length
            *    18767    18866: gap of unknown length
            *    18867    20034: contig of 1168 bp in length
            *    20035    20134: gap of unknown length
            *    20135    21336: contig of 1202 bp in length
            *    21337    21436: gap of unknown length
            *    21437    22545: contig of 1109 bp in length
            *    22546    22645: gap of unknown length
            *    22646    23755: contig of 1110 bp in length
            *    23756    23855: gap of unknown length
            *    23856    24927: contig of 1072 bp in length
            *    24928    25027: gap of unknown length
            *    25028    26153: contig of 1126 bp in length
            *    26154    26253: gap of unknown length
            *    26254    27455: contig of 1202 bp in length
            *    27456    27555: gap of unknown length
            *    27556    28679: contig of 1124 bp in length
            *    28680    28779: gap of unknown length
            *    28780    29893: contig of 1114 bp in length
            *    29894    29993: gap of unknown length
            *    29994    31172: contig of 1179 bp in length
            *    31173    31272: gap of unknown length
            *    31273    32388: contig of 1116 bp in length
            *    32389    32488: gap of unknown length
            *    32489    33650: contig of 1162 bp in length
            *    33651    33750: gap of unknown length
            *    33751    34987: contig of 1237 bp in length
            *    34988    35087: gap of unknown length
            *    35088    36153: contig of 1066 bp in length
            *    36154    36253: gap of unknown length
            *    36254    37421: contig of 1168 bp in length
            *    37422    37521: gap of unknown length
            *    37522    38717: contig of 1196 bp in length
            *    38718    38817: gap of unknown length
            *    38818    39915: contig of 1098 bp in length
            *    39916    40015: gap of unknown length
            *    40016    41215: contig of 1200 bp in length
            *    41216    41315: gap of unknown length
            *    41316    42350: contig of 1035 bp in length
            *    42351    42450: gap of unknown length
            *    42451    43607: contig of 1157 bp in length
            *    43608    43707: gap of unknown length
            *    43708    44880: contig of 1173 bp in length
            *    44881    44980: gap of unknown length
            *    44981    46095: contig of 1115 bp in length
            *    46096    46195: gap of unknown length
            *    46196    47463: contig of 1268 bp in length
            *    47464    47563: gap of unknown length
            *    47564    48771: contig of 1208 bp in length
            *    48772    48871: gap of unknown length
            *    48872    50000: contig of 1129 bp in length
            *    50001    50100: gap of unknown length
            *    50101    51241: contig of 1141 bp in length
            *    51242    51341: gap of unknown length
            *    51342    52439: contig of 1098 bp in length
            *    52440    52539: gap of unknown length
            *    52540    53719: contig of 1180 bp in length
            *    53720    53819: gap of unknown length
            *    53820    54972: contig of 1153 bp in length
            *    54973    55072: gap of unknown length
            *    55073    56256: contig of 1184 bp in length
            *    56257    56356: gap of unknown length
```

```
*    56357    57447: contig of 1091 bp in length
*    57448    57547: gap of unknown length
*    57548    58710: contig of 1163 bp in length
*    58711    58810: gap of unknown length
*    58811    59999: contig of 1189 bp in length
*    60000    60099: gap of unknown length
*    60100    61343: contig of 1244 bp in length
*    61344    61443: gap of unknown length
*    61444    62568: contig of 1125 bp in length
*    62569    62668: gap of unknown length
*    62669    63904: contig of 1236 bp in length
*    63905    64004: gap of unknown length
*    64005    65045: contig of 1041 bp in length
*    65046    65145: gap of unknown length
*    65146    66360: contig of 1215 bp in length
*    66361    66460: gap of unknown length
*    66461    67705: contig of 1245 bp in length
*    67706    67805: gap of unknown length
*    67806    68914: contig of 1109 bp in length
*    68915    69014: gap of unknown length
*    69015    70199: contig of 1185 bp in length
*    70200    70299: gap of unknown length
*    70300    71355: contig of 1056 bp in length
*    71356    71455: gap of unknown length
*    71456    73215: contig of 1760 bp in length
*    73216    73315: gap of unknown length
*    73316    74456: contig of 1141 bp in length
*    74457    74556: gap of unknown length
*    74557    76312: contig of 1756 bp in length
*    76313    76412: gap of unknown length
*    76413    77484: contig of 1072 bp in length
*    77485    77584: gap of unknown length
*    77585    78683: contig of 1099 bp in length
*    78684    78783: gap of unknown length
*    78784    79858: contig of 1075 bp in length
*    79859    79958: gap of unknown length
*    79959    80987: contig of 1029 bp in length
*    80988    81087: gap of unknown length
*    81088    82188: contig of 1101 bp in length
*    82189    82288: gap of unknown length
*    82289    83320: contig of 1032 bp in length
*    83321    83420: gap of unknown length
*    83421    84521: contig of 1101 bp in length
*    84522    84621: gap of unknown length
*    84622   101509: contig of 16888 bp in length.
FEATURES             Location/Qualifiers
     source          1. .101509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-167D21"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    27989 a   7533 c  28794 g   7012 t  30181 others
ORIGIN

  Query Match           14.3%;  Score 70.4;  DB 2;  Length 101509;
  Best Local Similarity  43.3%;  Pred. No. 0.002;
  Matches  143;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | |  ||||  |        |    | |||  |  || || ||  | |   |||    | |
Db   31049 CNCNNCCCCCCCCGCCCCCCNNCNCCCCCCCCCCCCCCCCCCCCCCCNNNCCCCNCCCNC 30990

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
            ||  |  || || || |||   |  ||    | | |  ||||  || | || || |
Db   30989 CCCCCCCCCCCCCCCCNCCCNNNCCCCCCNNCCCCNCNNCCCCCCCCCCCCCCCCNCNNNN 30930

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | |  |       |  || |    || | |||    |  || |   |
Db   30929 CCCCCCCCCCCCCCCCCNNCNNNNNNNCNCCCCCNNNCCCCNCCCCCCCCCCCCCCNNCC 30870

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
                 |      | |   |      |  |||   | |||  |   |      |||  |

Db   30869 CCCNNNCCCNCCNCNCNCNNNCNCCNCCGGCCGGCGCGNGCCGGGGCGGGGCGGCGGGGG 30810

Qy     257 GCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCG 316
           |||| |||| |  |  |  | |  |||| |  || |  |  | ||| |   | || || |
Db   30809 GCCGNCCGGGGGCCCCCCNCGGGCGGGGGGCGGGGGGGGCGGGGGCGGNCCGGGGGGGGG 30750

Qy     317 CAACGGGCGGCGGACGCTGCTGGGGAGGTG 346
               ||||        |  |   |  || |
Db   30749 NNCNGGGCCNNCCCNNCCCCNCNGCCGGGG 30720


RESULT 11
AC125906
LOCUS       AC125906               151358 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-77F8, *** SEQUENCING IN PROGRESS ***,
            63 unordered pieces.
ACCESSION   AC125906
VERSION     AC125906.1  GI:21671504
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 151358)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford J., Foster,P., Frantz,P.,
            Gabisi,A., Gao J., Garcia,A., Garner,T  Garza,N., Gill,R.,
            Gorrell,J.H.,  levara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151358)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

REFERENCE 3 (bases 1 to 151358)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GVRX
Center clone name: CH230-77F8
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 92473 bases at least Q40
Consensus quality: 98495 bases at least Q30
Consensus quality: 102505 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*        1     1032: contig of 1032 bp in length
*     1033     1132: gap of unknown length
*     1133     2134: contig of 1002 bp in length
*     2135     2234: gap of unknown length
*     2235     3255: contig of 1021 bp in length
*     3256     3355: gap of unknown length
*     3356     4398: contig of 1043 bp in length
*     4399     4498: gap of unknown length
*     4499     5857: contig of 1359 bp in length
*     5858     5957: gap of unknown length
*     5958     7206: contig of 1249 bp in length
*     7207     7306: gap of unknown length
*     7307     8379: contig of 1073 bp in length
*     8380     8479: gap of unknown length
*     8480     9982: contig of 1503 bp in length
*     9983    10082: gap of unknown length
*    10083    11373: contig of 1291 bp in length
*    11374    11473: gap of unknown length
*    11474    12776: contig of 1303 bp in length
*    12777    12876: gap of unknown length
*    12877    14025: contig of 1149 bp in length
*    14026    14125: gap of unknown length
*    14126    15578: contig of 1453 bp in length
*    15579    15678: gap of unknown length
*    15679    16957: contig of 1279 bp in length
*    16958    17057: gap of unknown length
*    17058    18241: contig of 1184 bp in length
*    18242    18341: gap of unknown length
*    18342    19373: contig of 1032 bp in length
*    19374    19473: gap of unknown length
*    19474    21050: contig of 1577 bp in length
*    21051    21150: gap of unknown length
*    21151    22862: contig of 1712 bp in length
*    22863    22962: gap of unknown length
*    22963    24563: contig of 1601 bp in length
*    24564    24663: gap of unknown length
*    24664    25683: contig of 1020 bp in length
*    25684    25783: gap of unknown length
*    25784    26831: contig of 1048 bp in length
*    26832    26931: gap of unknown length
*    26932    28243: contig of 1312 bp in length
*    28244    28343: gap of unknown length
*    28344    29758: contig of 1415 bp in length
*    29759    29858: gap of unknown length
*    29859    31174: contig of 1316 bp in length
*    31175    31274: gap of unknown length
*    31275    33433: contig of 2159 bp in length
*    33434    33533: gap of unknown length
*    33534    35127: contig of 1594 bp in length
*    35128    35227: gap of unknown length
*    35228    36393: contig of 1166 bp in length
*    36394    36493: gap of unknown length
*    36494    38614: contig of 2121 bp in length
*    38615    38714: gap of unknown length
*    38715    40665: contig of 1951 bp in length
*    40666    40765: gap of unknown length
*    40766    41885: contig of 1120 bp in length
*    41886    41985: gap of unknown length
*    41986    44086: contig of 2101 bp in length
*    44087    44186: gap of unknown length
*    44187    46007: contig of 1821 bp in length
*    46008    46107: gap of unknown length
*    46108    47865: contig of 1758 bp in length
*    47866    47965: gap of unknown length
*    47966    50021: contig of 2056 bp in length
*    50022    50121: gap of unknown length
*    50122    51983: contig of 1862 bp in length
*    51984    52083: gap of unknown length
*    52084    54300: contig of 2217 bp in length
*    54301    54400: gap of unknown length
*    54401    55910: contig of 1510 bp in length
*    55911    56010: gap of unknown length
*    56011    57542: contig of 1532 bp in length
*    57543    57642: gap of unknown length
*    57643    59295: contig of 1653 bp in length
*    59296    59395: gap of unknown length
*    59396    62393: contig of 2998 bp in length
*    62394    62493: gap of unknown length
*    62494    64229: contig of 1736 bp in length
*    64230    64329: gap of unknown length
*    64330    66089: contig of 1760 bp in length
*    66090    66189: gap of unknown length
*    66190    68909: contig of 2720 bp in length
*    68910    69009: gap of unknown length
*    69010    71491: contig of 2482 bp in length
*    71492    71591: gap of unknown length
*    71592    73355: contig of 1764 bp in length
*    73356    73455: gap of unknown length
*    73456    75777: contig of 2322 bp in length
*    75778    75877: gap of unknown length
*    75878    78392: contig of 2515 bp in length
*    78393    78492: gap of unknown length
*    78493    81849: contig of 3357 bp in length
*    81850    81949: gap of unknown length
*    81950    84202: contig of 2253 bp in length
*    84203    84302: gap of unknown length
*    84303    86896: contig of 2594 bp in length
*    86897    86996: gap of unknown length
*    86997    90467: contig of 3471 bp in length
*    90468    90567: gap of unknown length
*    90568    93764: contig of 3197 bp in length
*    93765    93864: gap of unknown length
*    93865    96842: contig of 2978 bp in length
*    96843    96942: gap of unknown length
*    96943    99989: contig of 3047 bp in length
*    99990   100089: gap of unknown length
*   100090   103155: contig of 3066 bp in length
*   103156   103255: gap of unknown length
```

Query Match 14.2%; Score 69.8; DB 2; Length 151358;
Best Local Similarity 51.9%; Pred. No. 0.0025;
Matches 149; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy      84 CCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACC 143
```

```
                 || || || ||| | |  ||  | | |||  ||||  || | || || ||  |||   ||
Db    2498 CCCCCCCCCCCCCCCCNNCCCNCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCACCCCCCCCCC 2557

Qy     144 CCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCG 203
             |   | | || || || || || | ||||| | ||| | || ||| |   |    |||
Db    2558 TCTTTTTCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCTCTCCTCCCCCCCCCCGGCG 2617

Qy     204 CGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCC 263
             |   ||| | |   |       | ||    |  || ||  | |||| | || ||  || |
Db    2618 GGCCCGGGGGGCGCGCTCNCCNTCCGCGCGGCCCCCCGGCCCCCGCGGCGGGGGGGGGGC 2677

Qy     264 GGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTCGTGGAGGTGGCGCAACGGG 323
             || | || || || |  | ||    || |       ||  |  || || || ||  |||
Db    2678 GGGGCGCGGCCGCGGGCGCGGGCGGGGGGCGCCGCGGGGGGCGGGGGGCGGGGCGGCGGC 2737

Qy     324 CGGCGGACGC1GCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGG 370
           |||  |  ||    |||   |  |    |  | |  || |  ||||
Db    2738 CGGGCGCGGCCCGCGGGCGCGCTCCGGCGGCCTGCCCCCCCGCGCGG 2784


RESULT 12
AC121227
LOCUS       AC121227               174605 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-103J5, *** SEQUENCING IN PROGRESS
            ***, 80 unordered pieces.
ACCESSION   AC121227
VERSION     AC121227.2  GI:21909120
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 174605)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Paxton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174605)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 174605)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 19, 2002 this sequence version replaced gi:20806257.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GYXK
                Center clone name: CH230-103J5
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 110654 bases at least Q40
                Consensus quality: 117178 bases at least Q30
                Consensus quality: 122597 bases at least Q20
            ------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 80 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1007: contig of 1007 bp in length
            *     1008     1107: gap of unknown length
            *     1108     2450: contig of 1343 bp in length
            *     2451     2550: gap of unknown length
            *     2551     4232: contig of 1682 bp in length
            *     4233     4332: gap of unknown length
            *     4333     5794: contig of 1462 bp in length
            *     5795     5894: gap of unknown length
            *     5895     6962: contig of 1068 bp in length
            *     6963     7062: gap of unknown length
            *     7063     8247: contig of 1185 bp in length
            *     8248     8347: gap of unknown length
            *     8348     9503: contig of 1156 bp in length
            *     9504     9603: gap of unknown length
            *     9604    10846: contig of 1243 bp in length
            *    10847    10946: gap of unknown length
            *    10947    12170: contig of 1224 bp in length
            *    12171    12270: gap of unknown length
            *    12271    13571: contig of 1301 bp in length
            *    13572    13671: gap of unknown length
            *    13672    14731: contig of 1060 bp in length
            *    14732    14831: gap of unknown length
            *    14832    16136: contig of 1305 bp in length
            *    16137    16236: gap of unknown length
            *    16237    17769: contig of 1533 bp in length
            *    17770    17869: gap of unknown length
            *    17870    19119: contig of 1250 bp in length
            *    19120    19219: gap of unknown length
            *    19220    20422: contig of 1203 bp in length
            *    20423    20522: gap of unknown length
            *    20523    22121: contig of 1599 bp in length
```

```
*    22122    22221: gap of unknown length
*    22222    24265: contig of 2044 bp in length
*    24266    24365: gap of unknown length
*    24366    25970: contig of 1605 bp in length
*    25971    26070: gap of unknown length
*    26071    27332: contig of 1262 bp in length
*    27333    27432: gap of unknown length
*    27433    28627: contig of 1195 bp in length
*    28628    28727: gap of unknown length
*    28728    29885: contig of 1158 bp in length
*    29886    29985: gap of unknown length
*    29986    31804: contig of 1819 bp in length
*    31805    31904: gap of unknown length
*    31905    33339: contig of 1435 bp in length
*    33340    33439: gap of unknown length
*    33440    35142: contig of 1703 bp in length
*    35143    35242: gap of unknown length
*    35243    36736: contig of 1494 bp in length
*    36737    36836: gap of unknown length
*    36837    38640: contig of 1804 bp in length
*    38641    38740: gap of unknown length
*    38741    39914: contig of 1174 bp in length
*    39915    40014: gap of unknown length
*    40015    41570: contig of 1556 bp in length
*    41571    41670: gap of unknown length
*    41671    43016: contig of 1346 bp in length
*    43017    43116: gap of unknown length
*    43117    44156: contig of 1040 bp in length
*    44157    44256: gap of unknown length
*    44257    45989: contig of 1733 bp in length
*    45990    46089: gap of unknown length
*    46090    47638: contig of 1549 bp in length
*    47639    47738: gap of unknown length
*    47739    49218: contig of 1480 bp in length
*    49219    49318: gap of unknown length
*    49319    50429: contig of 1111 bp in length
*    50430    50529: gap of unknown length
*    50530    52925: contig of 2396 bp in length
*    52926    53025: gap of unknown length
*    53026    54637: contig of 1612 bp in length
*    54638    54737: gap of unknown length
*    54738    57055: contig of 2318 bp in length
*    57056    57155: gap of unknown length
*    57156    58371: contig of 1216 bp in length
*    58372    58471: gap of unknown length
*    58472    60149: contig of 1678 bp in length
*    60150    60249: gap of unknown length
*    60250    61548: contig of 1299 bp in length
*    61549    61648: gap of unknown length
*    61649    63896: contig of 2248 bp in length
*    63897    63996: gap of unknown length
*    63997    66103: contig of 2107 bp in length
*    66104    66203: gap of unknown length
*    66204    68167: contig of 1964 bp in length
*    68168    68267: gap of unknown length
*    68268    69807: contig of 1540 bp in length
*    69808    69907: gap of unknown length
*    69908    71931: contig of 2024 bp in length
*    71932    72031: gap of unknown length
*    72032    73191: contig of 1160 bp in length
*    73192    73291: gap of unknown length
*    73292    74767: contig of 1476 bp in length
*    74768    74867: gap of unknown length
*    74868    76389: contig of 1522 bp in length
*    76390    76489: gap of unknown length
*    76490    77915: contig of 1426 bp in length
*    77916    78015: gap of unknown length
*    78016    80236: contig of 2221 bp in length
*    80237    80336: gap of unknown length
*    80337    82748: contig of 2412 bp in length
*    82749    82848: gap of unknown length
*    82849    85565: contig of 2717 bp in length
*    85566    85665: gap of unknown length
*    85666    87488: contig of 1823 bp in length
*    87489    87588: gap of unknown length
*    87589    88625: contig of 1037 bp in length

  Query Match             13.9%;  Score 68.4;  DB 2;  Length 174605;
  Best Local Similarity   49.2%;  Pred. No. 0.0044;
  Matches  150;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy      3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
          | ||   |  | | ||||                |     |||  |  || || ||| |
Db  17159 CCCCCTCCCCCTTCCTCCNNNNNNNNNNNNNCCTNCCCNNNCCCCCCCCCCCCCCCCCGCCC 17218

Qy     63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
          || ||||   || | || |  || || || ||| | |  || || | |||  |||| ||
Db  17219 CCCCCCCGCCGCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCC 17278

Qy    123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182
          | || || |   |||    |||| | | | |  || || |  ||  | || |  |||||
Db  17279 CCCCCNCCCCCCCCCCCCCGCCCCCCCCCCCCCGCCCCCCCCNCCCCGGGCGGCGGGCCCG 17338

Qy    183 TGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGT 242
          | || |          | ||||| |       |   |||  | |   ||   | |||
Db  17339 GGGCGGCGCGCCGGGGGGGCGCGGGCCCCGCCGCCCCGGCGCGCGGGGGGCGCGCCGGCG 17398

Qy    243 TCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGC 302
          ||||||   | |||     ||  | |  ||||  | | ||  |  |||  ||   || |
Db  17399 CCGCCGCCCCGCCGGGGCCGCGCGGGGGCGCGGGCGCGCGGCGGGCGGCCGCGCGCCGCC 17458

Qy    303 TGGTG 307
          || |
Db  17459 GGGCG 17463


RESULT 13
AC118303
LOCUS       AC118303               174031 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-173H17, *** SEQUENCING IN PROGRESS
            ***, 58 unordered pieces.
ACCESSION   AC118303
VERSION     AC118303.5  GI:21746412
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 174031)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D , Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
```

Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
  TITLE Direct Submission
  JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174031)
  AUTHORS Worley,K.C.
  TITLE Direct Submission
  JOURNAL Submitted (15-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 174031)
  AUTHORS Worley,K.C.
  TITLE Direct Submission
  JOURNAL Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 14, 2002 this sequence version replaced gi:20330629.
    ---------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
    ---------- Project Information
    Center project name: GUJG
    Center clone name: CH230-173H17
    ---------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 107398 bases at least Q40
    Consensus quality: 119658 bases at least Q30
    Consensus quality: 127269 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1024: contig of 1024 bp in length
* 1025 1124: gap of unknown length
* 1125 2167: contig of 1043 bp in length
* 2168 2267: gap of unknown length
* 2268 3414: contig of 1147 bp in length
* 3415 3514: gap of unknown length
* 3515 4714: contig of 1200 bp in length
* 4715 4814: gap of unknown length
* 4815 6353: contig of 1539 bp in length
* 6354 6453: gap of unknown length
* 6454 7915: contig of 1462 bp in length
* 7916 8015: gap of unknown length
* 8016 9359: contig of 1344 bp in length
* 9360 9459: gap of unknown length
* 9460 11082: contig of 1623 bp in length
* 11083 11182: gap of unknown length
* 11183 12369: contig of 1187 bp in length
* 12370 12469: gap of unknown length
* 12470 13930: contig of 1461 bp in length
* 13931 14030: gap of unknown length
* 14031 15253: contig of 1223 bp in length
* 15254 15353: gap of unknown length
* 15354 16696: contig of 1343 bp in length
* 16697 16796: gap of unknown length
* 16797 18403: contig of 1607 bp in length
* 18404 18503: gap of unknown length
* 18504 19776: contig of 1273 bp in length
* 19777 19876: gap of unknown length
* 19877 21648: contig of 1772 bp in length
* 21649 21748: gap of unknown length
* 21749 23094: contig of 1346 bp in length
* 23095 23194: gap of unknown length
* 23195 25092: contig of 1898 bp in length
* 25093 25192: gap of unknown length
* 25193 26802: contig of 1610 bp in length
* 26803 26902: gap of unknown length
* 26903 28726: contig of 1824 bp in length
* 28727 28826: gap of unknown length
* 28827 31064: contig of 2238 bp in length
* 31065 31164: gap of unknown length
* 31165 33254: contig of 2090 bp in length
* 33255 33354: gap of unknown length
* 33355 35763: contig of 2409 bp in length
* 35764 35863: gap of unknown length
* 35864 36989: contig of 1126 bp in length
* 36990 37089: gap of unknown length
* 37090 38703: contig of 1614 bp in length
* 38704 38803: gap of unknown length
* 38804 40934: contig of 2131 bp in length
* 40935 41034: gap of unknown length
* 41035 42978: contig of 1944 bp in length
* 42979 43078: gap of unknown length
* 43079 45214: contig of 2136 bp in length
* 45215 45314: gap of unknown length
* 45315 47780: contig of 2466 bp in length
* 47781 47880: gap of unknown length
* 47881 50209: contig of 2329 bp in length
* 50210 50309: gap of unknown length
* 50310 52364: contig of 2055 bp in length
* 52365 52464: gap of unknown length
* 52465 54739: contig of 2275 bp in length
* 54740 54839: gap of unknown length
* 54840 56920: contig of 2081 bp in length
* 56921 57020: gap of unknown length
* 57021 59676: contig of 2656 bp in length
* 59677 59776: gap of unknown length
* 59777 62058: contig of 2282 bp in length
* 62059 62158: gap of unknown length
* 62159 64481: contig of 2323 bp in length
* 64482 64581: gap of unknown length
* 64582 67737: contig of 3156 bp in length
* 67738 67837: gap of unknown length
* 67838 69856: contig of 2019 bp in length
* 69857 69956: gap of unknown length
* 69957 72698: contig of 2742 bp in length
* 72699 72798: gap of unknown length
* 72799 75900: contig of 3102 bp in length
* 75901 76000: gap of unknown length
* 76001 78614: contig of 2614 bp in length
* 78615 78714: gap of unknown length
* 78715 81699: contig of 2985 bp in length
* 81700 81799: gap of unknown length
* 81800 85909: contig of 4110 bp in length
* 85910 86009: gap of unknown length
* 86010 89796: contig of 3787 bp in length
* 89797 89896: gap of unknown length
* 89897 93635: contig of 3739 bp in length
* 93636 93735: gap of unknown length
* 93736 97658: contig of 3923 bp in length
* 97659 97758: gap of unknown length
* 97759 101045: contig of 3287 bp in length

```
            *   101046   101145: gap of unknown length
            *   101146   105352: contig of 4207 bp in length
            *   105353   105452: gap of unknown length
            *   105453   109545: contig of 4093 bp in length
            *   109546   109645: gap of unknown length
            *   109646   114476: contig of 4831 bp in length
            *   114477   114576: gap of unknown length
            *   114577   118745: contig of 4169 bp in length
            *   118746   118845: gap of unknown length
            *   118846   122388: contig of 3543 bp in length
            *   122389   122488: gap of unknown length
            *   122489   128899: contig of 6411 bp in length
            *   128900   128999: gap of unknown length
            *   129000   134742: contig of 5743 bp in length
            *   134743   134842: gap of unknown length
            *   134843   139853: contig of 5011 bp in length

 Query Match           13.8%;  Score 68;  DB 2;  Length 174031;
 Best Local Similarity  49.5%;  Pred. No. 0.0052;
 Matches  158;  Conservative    0;  Mismatches  159;  Indels    2;  Gaps    1;

 Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
            | ||  |||  |       ||     ||      || || ||    || | ||    | |
 Db   24645 CCCCCNCCCCCCCNCCCCCCCCCNNCNNCCNNCNCCCCCCCCCCCNCGCCCCCNCCCCCCCNC 24704

 Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTCCCCCGGCCTCGCCTCCGCAAAC 136
              ||    || || || ||| | |  || || | |||  ||||  || | ||       |
 Db   24705 CCCCNCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNNNNNCNC 24764

 Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
            ||   ||||| | | |  | || || ||  | |     | || | | | || | | |
 Db   24765 CCNCCCCCCCCCGCCCCNCCCCCCCCCCCGCCCCNNNNCCCGGCGGCGGCGGCGGGGCNG 24824

 Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
              | ||| | | ||| ||  |  |  | |   ||   ||||  ||   |  |  ||   |
 Db   24825 CGGGGCGGGGGNGGGGGCGGGGGGCGGCGGCGGGGGGGCGCGCCGGGGGGGGGGGGCCCG 24884

 Qy     257 GCCG--GCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGG 314
             |||  ||||| |  | |||   | ||||| |  ||||  ||    || || |   |  |
 Db   24885 CCCGGCGCCGGGGGCCGGCGCGCGGGGGGGCGGGGGCGGGGGGCGCGCGGGCGCNCGGCG 24944

 Qy     315 CGCAACGGGCGGCGGACGC 333
            |||    |||||||  | |
 Db   24945 CGCGGGCGGCGGCGCGCCC 24963


RESULT 14
AC106398
 OCUS       AC106398               68787 bp    DNA     linear   HTG 13-JUL-2002
 EFINITION  Rattus norvegicus clone CH230-137B1, *** SEQUENCING IN PROGRESS
            ***, 37 unordered pieces.
ACCESSION   AC106398
VERSION     AC106398.2  GI:21735179
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 68787)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68787)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 68787)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18138919.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GKUW
                Center clone name: CH230-137B1
            ---------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 35643 bases at least Q40
                Consensus quality: 39014 bases at least Q30
                Consensus quality: 41785 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1070: contig of 1070 bp in length
            *     1071     1170: gap of unknown length
            *     1171     2678: contig of 1508 bp in length
            *     2679     2778: gap of unknown length
            *     2779     4068: contig of 1290 bp in length
```

```
*     4069     4168: gap of unknown length
*     4169     5608: contig of 1440 bp in length
*     5609     5708: gap of unknown length
*     5709     7127: contig of 1419 bp in length
*     7128     7227: gap of unknown length
*     7228     8577: contig of 1350 bp in length
*     8578     8677: gap of unknown length
*     8678     9793: contig of 1116 bp in length
*     9794     9893: gap of unknown length
*     9894    10907: contig of 1014 bp in length
*    10908    11007: gap of unknown length
*    11008    12349: contig of 1342 bp in length
*    12350    12449: gap of unknown length
*    12450    13548: contig of 1099 bp in length
*    13549    13648: gap of unknown length
*    13649    14730: contig of 1082 bp in length
*    14731    14830: gap of unknown length
*    14831    16251: contig of 1421 bp in length
*    16252    16351: gap of unknown length
*    16352    18478: contig of 2127 bp in length
*    18479    18578: gap of unknown length
*    18579    20188: contig of 1610 bp in length
*    20189    20288: gap of unknown length
*    20289    21551: contig of 1263 bp in length
*    21552    21651: gap of unknown length
*    21652    23305: contig of 1654 bp in length
*    23306    23405: gap of unknown length
*    23406    25560: contig of 2155 bp in length
*    25561    25660: gap of unknown length
*    25661    27283: contig of 1623 bp in length
*    27284    27383: gap of unknown length
*    27384    29246: contig of 1863 bp in length
*    29247    29346: gap of unknown length
*    29347    31576: contig of 2230 bp in length
*    31577    31676: gap of unknown length
*    31677    33313: contig of 1637 bp in length
*    33314    33413: gap of unknown length
*    33414    35876: contig of 2463 bp in length
*    35877    35976: gap of unknown length
*    35977    37553: contig of 1577 bp in length
*    37554    37653: gap of unknown length
*    37654    39176: contig of 1523 bp in length
*    39177    39276: gap of unknown length
*    39277    40741: contig of 1465 bp in length
*    40742    40841: gap of unknown length
*    40842    42390: contig of 1549 bp in length
*    42391    42490: gap of unknown length
*    42491    44654: contig of 2164 bp in length
*    44655    44754: gap of unknown length
*    44755    46171: contig of 1417 bp in length
*    46172    46271: gap of unknown length
*    46272    48302: contig of 2031 bp in length
*    48303    48402: gap of unknown length
*    48403    50074: contig of 1672 bp in length
*    50075    50174: gap of unknown length
*    50175    52341: contig of 2167 bp in length
*    52342    52441: gap of unknown length
*    52442    54601: contig of 2160 bp in length
*    54602    54701: gap of unknown length
*    54702    57006: contig of 2305 bp in length
*    57007    57106: gap of unknown length
*    57107    59715: contig of 2609 bp in length
*    59716    59815: gap of unknown length
*    59816    62344: contig of 2529 bp in length
*    62345    62444: gap of unknown length
*    62445    66001: contig of 3557 bp in length
*    66002    66101: gap of unknown length
*    66102    68787: contig of 2686 bp in length.
FEATURES             Location/Qualifiers
     source          1. .68787
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-137B1"
```

```
BASE COUNT    18229 a  14391 c  13550 g  17475 t   5142 others
ORIGIN

  Query Match             13.7%;  Score 67.4;  DB 2;  Length 68787;
  Best Local Similarity   54.1%;  Pred. No. 0.0071;
  Matches  157;  Conservative    0;  Mismatches  132;  Indels    1;  Gaps    1;

Qy      16 TCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCT 75
           ||  | ||||  |  |       | |||||  |  || || ||  | || ||||  | ||
Db   31191 TCCTCNCCCCCCCCCTGCTCGTC'CCTC'CCCCCCCCCCCCCCCTCCCCCCCCCCCCTCCT 31250

Qy      76 CTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAA 135
           | || || || || || ||| | |   | || | |||||||||  || | || || |
Db   31251 CCCCCCTCCCCCCCCCCCCCCCCCCCCGCTCCCCCCCCCTTCCCCCCCCCCCCCCCCTCTCC 31310

Qy     136 CCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGT 195
           |||   ||||| | | | |    ||| || || |  |||     | | | ||  | ||
Db   31311 CCCCCCCCCCCTCCCCCCCTCGTTCCCCCCCCCCCCGCGCGGTCTCGGCGGCGCGCCCGCG 31370

Qy     196 CTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGC 255
              ||||   | ||||| ||||  |     ||     ||| ||   |  ||||  ||
Db   31371 GCGCGTCGCGGCGGCGCGCCCGCCGGCGGCCGGGGCCGCCGCCGCG-GGGGGGCTGCCGC 31429

Qy     256 GGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGG 305
           ||||||  |    ||    || ||||||     |||| ||  |||   ||
Db   31430 CGCCGGCGGCGCCGCCTGCGCCGTGGGGCGTCGGGCGGCGCTGCGCGGGG 31479


RESULT 15
AC126730
LOCUS       AC126730               130899 bp    DNA     linear   HTG 23-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-6H4, *** SEQUENCING IN PROGRESS ***,
            69 unordered pieces.
ACCESSION   AC126730
VERSION     AC126730.2  GI:21886838
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 130899)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
```

Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 130899)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 130899)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 17, 2002 this sequence version replaced gi:21716710.
---------- Genome Center
Center: Baylor College of Medic ne
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GBWP
Center clone name: CH230-6H4
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 71039 bases at least Q40
Consensus quality: 77045 bases at least Q30
Consensus quality: 80907 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*        1     1027: contig of 1027 bp in length
*     1028     1127: gap of unknown length
*     1128     2272: contig of 1145 bp in length
*     2273     2372: gap of unknown length
*     2373     3398: contig of 1026 bp in length
*     3399     3498: gap of unknown length
*     3499     4802: contig of 1304 bp in length
*     4803     4902: gap of unknown length
*     4903     5993: contig of 1091 bp in length
*     5994     6093: gap of unknown length
*     6094     7244: contig of 1151 bp in length
*     7245     7344: gap of unknown length
*     7345     8880: contig of 1536 bp in length
*     8881     8980: gap of unknown length
*     8981     9987: contig of 1007 bp in length
*     9988    10087: gap of unknown length
*    10088    11342: contig of 1255 bp in length
*    11343    11442: gap of unknown length
*    11443    12560: contig of 1118 bp in length
*    12561    12660: gap of unknown length
*    12661    13884: contig of 1224 bp in length
*    13885    13984: gap of unknown length
*    13985    15064: contig of 1080 bp in length
*    15065    15164: gap of unknown length
*    15165    16546: contig of 1382 bp in length
*    16547     6646: gap of unknown length
*    16647    18513: contig of 1867 bp in length
*    18514    18613: gap of unknown length
*    18614    19705: contig of 1092 bp in length
*    19706    19805: gap of unknown length
*    19806    21270: contig of 1465 bp in length
*    21271    21370: gap of unknown length
*    21371    22600: contig of 1230 bp in length
*    22601    22700: gap of unknown length
*    22701    23741: contig of 1041 bp in length
*    23742    23841: gap of unknown length
*    23842    24945: contig of 1104 bp in length
*    24946    25045: gap of unknown length
*    25046    27258: contig of 2213 bp in length
*    27259    27358: gap of unknown length
*    27359     8617: contig of 1259 bp in length
*    28618    28717: gap of unknown length
*    28718    29892: contig of 1175 bp in length
*    29893    29992: gap of unknown length
*    29993    31396: contig of 1404 bp in length
*    31397    31496: gap of unknown length
*    31497    32504: contig of 1008 bp in length
*    32505    32604: gap of unknown length
*    32605    34000: contig of 1396 bp in length
*    34001    34100: gap of unknown length
*    34101    35857: contig of 1757 bp in length
*    35858    35957: gap of unknown length
*    35958    37377: contig of 1420 bp in length
*    37378    37477: gap of unknown length
*    37478    38497: contig of 1020 bp in length
*    38498    38597: gap of unknown length
*    38598    39693: contig of 1096 bp in length
*    39694    39793: gap of unknown length
*    39794    42502: contig of 2709 bp in length
*    42503    42602: gap of unknown length
*    42603    44423: contig of 1821 bp in length
*    44424    44523: gap of unknown length
*    44524    46230: contig of 1707 bp in length
*    46231    46330: gap of unknown length
*    46331    47759: contig of 1429 bp in length
*    47760    47859: gap of unknown length
*    47860    49208: contig of 1349 bp in length
*    49209    49308: gap of unknown length
*    49309    50360: contig of 1052 bp in length
*    50361    50460: gap of unknown length
*    50461    52257: contig of 1797 bp in length
*    52258    52357: gap of unknown length
*    52358    54147: contig of 1790 bp in length
*    54148    54247: gap of unknown length
*    54248    55402: contig of 1155 bp in length
*    55403    55502: gap of unknown length
*    55503    57008: contig of 1506 bp in length
*    57009    57108: gap of unknown length
*    57109    58547: contig of 1439 bp in length
*    58548    58647: gap of unknown length
*    58648    60159: contig of 1512 bp in length
*    60160    60259: gap of unknown length
*    60260    61728: contig of 1469 bp in length
*    61729    61828: gap of unknown length
*    61829    64395: contig of 2567 bp in length
*    64396    64495: gap of unknown length
*    64496    66096: contig of 1601 bp in length
*    66097    66196: gap of unknown length
*    66197    67712: contig of 1516 bp in length
*    67713    67812: gap of unknown length
*    67813    68997: contig of 1185 bp in length
*    68998    69097: gap of unknown length
*    69098    70867: contig of 1770 bp in length
*    70868    70967: gap of unknown length
*    70968    72129: contig of 1162 bp in length
*    72130    72229: gap of unknown length
```

```
          *    72230    75006: contig of 2777 bp in length
          *    75007    75106: gap of unknown length
          *    75107    76249: contig of 1143 bp in length
          *    76250    76349: gap of unknown length
          *    76350    79053: contig of 2704 bp in length
          *    79054    79153: gap of unknown length
          *    79154    81129: contig of 1976 bp in length
          *    81130    81229: gap of unknown length
          *    81230    83849: contig of 2620 bp in length
          *    83850    83949: gap of unknown length
          *    83950    86640: contig of 2691 bp in length

 Query Match             13.7%;  Score 67.4;  DB 2;  Length 130899;
 Best Local Similarity   46.3%;  Pred. No. 0.0068;
 Matches  158;  Conservative    0;  Mismatches  183;  Indels    0;  Gaps    0;

Qy      36 CCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCC 95
           ||   | |||  |  || || ||    || ||||        || |  ||  | || |||
Db   89941 CCCCCCCCCCCCCNCCCNCCCCCNNNCCCCCCCCCCCCCNNNNCCCCCCCCCNCCCCCCCC 90000

Qy      96 ACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCC 155
             |  || | |||  |||| |    || || |   |||   ||||| | | | ||
Db   90001 NAAGCCCNCCCCCCCCCCCCCCCCCCNNNCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCC 90060

Qy     156 TCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCC 215
            || ||| |  |       | || |      |   || | |  |   | |   ||| |
Db   90061 CCCCCCGGCGACNNNNGCGCCCGNCGCGCCGGCGNGCCGGCGCGCCGGGGGCGGGGGGGG 90120

Qy     216 AGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGG 275
            ||    | |   ||   ||| | ||   ||||| |||   | |||||| | |  |
Db   90121 GGCGCCGGGGGCGGGGGGGGGCGCGCGGGCCGCGGGGNGCGGCGGCCGGGGCGGCGGCC 90180

Qy     276 CAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTG 335
           | |  ||   |  |||| | | ||||   | |  |  ||     ||| ||  |   |
Db   90181 CCGGCGGNCCGCGGGCGNCCGCGCGGGNCGNGCGGCGGGGCGCGCGGNCGCGGNCGGGGC 90240

Qy     336 CTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTGAG 376
           |    | || |  |  |  |  |     |  |||||  | |
Db   90241 CCCCCGCGGCGGGCGGGNGGCGCGCGGNCGGGCGGGGGGGG 90281

Search completed: November 19, 2002, 01:30:53
Job time : 3740 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 17, 2002, 14:17:57 ; Search time 2693 Seconds
(without alignments)
5316.963 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agctccgagcgtcattctcc.....agccaattttggngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	112	22.8	131704	2	AP004313	AP004313 Oryza sat
	2	112	22.8	161809	2	AP005173	AP005173 Oryza sat
	3	84.2	17.1	163381	2	AC128230	AC128230 Rattus no
	4	78	15.9	177168	2	AC107145	AC107145 Rattus no
	5	76.8	15.6	138467	2	AC111697	AC111697 Rattus no
	6	74.4	15.1	82815	2	AC062001	AC062001 Homo sapi
	7	73.2	14 9	158897	2	AC121439	AC121439 Rattus no
	8	72.2	14.7	68330	2	AC116109	AC116109 Mus muscu
	9	72.2	14.7	117082	2	AC126076	AC126076 Rattus no
c	10	70.4	14.3	101509	2	AC027353	AC027353 Homo sapi
	11	69.8	14.2	151358	2	AC125906	AC125906 Rattus no
	12	68.4	13.9	174605	2	AC121227	AC121227 Rattus no
	13	68	13.8	174031	2	AC118303	AC118303 Rattus no
	14	67.4	13.7	68787	2	AC106398	AC106398 Rattus no
	15	67.4	13.7	130899	2	AC126730	AC126730 Rattus no
	16	67.4	13.7	165988	2	AC121212	AC121212 Rattus no
	17	67.2	13.7	185263	2	AC129765	AC129765 Rattus no
	18	67.2	13.7	188231	2	AC094753	AC094753 Rattus no
	19	66.8	13.6	179608	2	AC128497	AC128497 Rattus no
	20	66.8	13.6	209216	2	AC117126	AC117126 Rattus no
c	21	66.6	13.5	99849	2	AC096824	AC096824 Rattus no
c	22	66.6	13.5	117082	2	AC126076	AC126076 Rattus no
c	23	66.6	13.5	118276	2	AP004120	AP004120 Oryza sat
	24	66.6	13.5	158117	2	AP005111	AP005111 Oryza sat
c	25	66.6	13.5	158971	2	AP004882	AP004882 Oryza sat
c	26	66.6	13.5	165988	2	AC121212	AC121212 Rattus no
	27	66.6	13.5	171624	2	AC095621	AC095621 Rattus no
	28	66.2	13.5	194776	2	AC129706	AC129706 Rattus no
	29	65.8	13.4	171204	2	AC130936	AC130936 Rattus no
	30	65.6	13.3	175041	2	AC097133	AC097133 Rattus no
	31	65.2	13.3	62839	2	AC098999	AC098999 Rattus no
	32	65.2	13.3	170696	2	AP004658	AP004658 Oryza sat
c	33	65.2	13.3	171204	2	AC130936	AC130936 Rattus no
c	34	65	13.2	184245	2	AC127355	AC127355 Mus muscu
	35	64.6	13.1	110000	2	AC116301_0	AC116301 Homo sapi
	36	64.4	13.1	171852	2	AC058782	AC058782 Homo sapi
	37	64.4	13.1	181022	2	AC121905	AC121905 Mus muscu
	38	64.4	13.1	188133	2	AC112428	AC112428 Rattus no
	39	64.2	13.0	81417	2	AC023265	AC023265 Homo sapi
	40	64	13.0	69417	2	AC121516	AC121516 Mus muscu
	41	64	13.0	162183	2	AC097753	AC097753 Rattus no
	42	63.8	13.0	161163	2	AC115099	AC115099 Homo sapi
	43	63.8	13.0	178941	2	AC113635	AC113635 Rattus no
c	44	63.6	12.9	129667	2	AC095300	AC095300 Rattus no
c	45	63.6	12.9	135119	2	AC011578	AC011578 Homo sapi

ALIGNMENTS

RESULT 1
AP004313/c
LOCUS AP004313 131704 bp DNA linear HTG 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 7 clone P0440B02,
*** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AP004313
VERSION AP004313.1 GI:16874543
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:P0440B02.
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1

```
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
            clone:P0440B02
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 131704)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .131704
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="7"
                     /clone="P0440B02"
BASE COUNT    38468 a  27101 c  28237 g  37798 t    100 others
ORIGIN

  Query Match             22.8%;  Score 112;  DB 2;  Length 131704;
  Best Local Similarity   60.1%;  Pred. No. 4.5e-11;
  Matches  206;  Conservative    0;  Mismatches  131;  Indels    6;  Gaps    1;

Qy      66 CCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCG 125
           ||| ||||||| |  |  ||||| ||  ||   |  || |  | |    || |  || ||
Db 108777 CCCAAATGCTCCCGGCCACCTCCACCGTCCCACCCACCTCGCCCCTTCCCCACTCCCGCG 108718

Qy     126 CCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGT 185
           |  || |   ||   ||| |||| |  | || |||| |     ||  ||  || | | |
Db 108717 CTGCCCCGCGCCGCCACCTCCGCGCAGGGCTTCTCCACGGGTCCGTGGCGACGGCGGCGG 108658

Qy     186 CGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCG 245
           |     ||  |   | | ||    | |||| |    ||| | ||    ||    |
Db 108657 TGGTTCGGGCGTGCGGGGTGACGACGGGCCACCGTCGGAGGATGGGCTCGGTTCGGGCGA 108598

 y     246 CCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGG------GGAGTTGGCGACGGAGC 299
           | |  |||| | |||  ||   |  |||   | ||       || | | | |  |||||
Db 108597 GCCCCAGGGCAGGCGGATGGGAGGTGGCGATGGCGGATAAGGAGGGGGTAGGGGTGGAGC 108538

Qy     300 GGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACT 359
           |||| ||||||||||||||  ||||||| ||||| || ||||||||||||| ||||||||
Db 108537 GGCTAGTGGAGGTGGCGCAGAGGGCGGCAGACGCCGCAGGGGAGGTGCTCAGGAAGTACT 108478

Qy     360 TCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCC 402
           |||| || || || |||||||||||||| ||||| || ||| |
Db 108477 TCCGGCAACGCGTCGAGATCATCGACAAGGAGGATCAGAGTGC 108435


RESULT 2
AP005173
LOCUS       AP005173               161809 bp    DNA     linear   HTG 17-MAY-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 7 clone
            OSJNBb0003E08, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AP005173
VERSION     AP005173.1  GI:20975299
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
            clone:OSJNBb0003E08.
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OSJNBb0003E08
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 161809)
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .161809
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="7"
                     /clone="OSJNBb0003E08"
BASE COUNT    46758 a  34068 c  34038 g  46734 t    211 others
ORIGIN

  Query Match             22.8%;  Score 112;  DB 2;  Length 161809;
  Best Local Similarity   60.1%;  Pred. No. 4.4e-11;
  Matches  206;  Conservative    0;  Mismatches  131;  Indels    6;  Gaps    1;

Qy      66 CCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCG 125
           ||| ||||||| |  |  ||||| ||  ||   |  || |  | |    || |  || ||
Db 108667 CCCAAATGCTCCCGGCCACCTCCACCGTCCCACCCACCTCGCCCCTTCCCCACTCCCGCG 108726

Qy     126 CCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGT 185
           |  || |   ||   ||| |||| |  | || |||| |     ||  ||  || | | |
Db 108727 CTGCCCCGCGCCGCCACCTCCGCGCAGGGCTTCTCCACGGGTCCGTGGCGACGGCGGCGG 108786

Qy     186 CGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCG 245
           |     ||  |   | | ||    | |||| |    ||| | ||    ||    |
Db 108787 TGGTTCGGGCGTGCGGGGTGACGACGGGCCACCGTCGGAGGATGGGCTCGGTTCGGGCGA 108846

Qy     246 CCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGG------GGAGTTGGCGACGGAGC 299
           | |  |||| | |||  ||   |  |||   | ||       || | | | |  |||||
Db 108847 GCCCCAGGGCAGGCGGATGGGAGGTGGCGATGGCGGATAAGGAGGGGGTAGGGGTGGAGC 108906

Qy     300 GGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACT 359
           |||| ||||||||||||||  ||||||| ||||| || ||||||||||||| ||||||||
Db 108907 GGCTAGTGGAGGTGGCGCAGAGGGCGGCAGACGCCGCAGGGGAGGTGCTCAGGAAGTACT 108966

Qy     360 TCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCC 402
           |||| || || || |||||||||||||| ||||| || ||| |
Db 108967 TCCGGCAACGCGTCGAGATCATCGACAAGGAGGATCAGAGTGC 109009


RESULT 3
AC128230
LOCUS       AC128230               163381 bp    DNA     linear   HTG 19-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-57K8, *** SEQUENCING IN PROGRESS ***,
            85 unordered pieces.
ACCESSION   AC128230
VERSION     AC128230.1  GI:21908840
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
```

ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 163381)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Joliv:t,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163381)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GYBT
Center clone name: CH230-57K8
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 78315 bases at least Q40
Consensus quality: 87206 bases at least Q30
Consensus quality: 92805 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 85 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1027: contig of 1027 bp in length
* 1028 1127: gap of unknown length
* 1128 2195: contig of 1068 bp in length
* 2196 2295: gap of unknown length
* 2296 3623: contig of 1328 bp in length
* 3624 3723: gap of unknown length
* 3724 5144: contig of 1421 bp in length
* 5145 5244: gap of unknown length
* 5245 6262: contig of 1018 bp in length
* 6263 6362: gap of unknown length
* 6363 7886: contig of 1524 bp in length
* 7887 7986: gap of unknown length
* 7987 9581: contig of 1595 bp in length
* 9582 9681: gap of unknown length
* 9682 11192: contig of 1511 bp in length
* 11193 11292: gap of unknown length
* 11293 12398: contig of 1106 bp in length
* 12399 12498: gap of unknown length
* 12499 13696: contig of 1198 bp in length
* 13697 13796: gap of unknown length
* 13797 14976: contig of 1180 bp in length
* 14977 15076: gap of unknown length
* 15077 16562: contig of 1486 bp in length
* 16563 16662: gap of unknown length
* 16663 17680: contig of 1018 bp in length
* 17681 17780: gap of unknown length
* 17781 18800: contig of 1020 bp in length
* 18801 18900: gap of unknown length
* 18901 20003: contig of 1103 bp in length
* 20004 20103: gap of unknown length
* 20104 21681: contig of 1578 bp in length
* 21682 21781: gap of unknown length
* 21782 22870: contig of 1089 bp in length
* 22871 22970: gap of unknown length
* 22971 24090: contig of 1120 bp in length
* 24091 24190: gap of unknown length
* 24191 25426: contig of 1236 bp in length
* 25427 25526: gap of unknown length
* 25527 26689: contig of 1163 bp in length
* 26690 26789: gap of unknown length
* 26790 28356: contig of 1567 bp in length
* 28357 28456: gap of unknown length
* 28457 30062: contig of 1606 bp in length
* 30063 30162: gap of unknown length
* 30163 31306: contig of 1144 bp in length
* 31307 31406: gap of unknown length
* 31407 32968: contig of 1562 bp in length
* 32969 33068: gap of unknown length
* 33069 34291: contig of 1223 bp in length
* 34292 34391: gap of unknown length
* 34392 35467: contig of 1076 bp in length
* 35468 35567: gap of unknown length
* 35568 36788: contig of 1221 bp in length
* 36789 36888: gap of unknown length
* 36889 38584: contig of 1696 bp in length
* 38585 38684: gap of unknown length
* 38685 40070: contig of 1386 bp in length
* 40071 40170: gap of unknown length
* 40171 41693: contig of 1523 bp in length
* 41694 41793: gap of unknown length
* 41794 42956: contig of 1163 bp in length
* 42957 43056: gap of unknown length
* 43057 44652: contig of 1596 bp in length
* 44653 44752: gap of unknown length
* 44753 46319: contig of 1567 bp in length
* 46320 46419: gap of unknown length
* 46420 48471: contig of 2052 bp in length

```
*    48472    48571: gap of unknown length
*    48572    49973: contig of 1402 bp in length
*    49974    50073: gap of unknown length
*    50074    51893: contig of 1820 bp in length
*    51894    51993: gap of unknown length
*    51994    53016: contig of 1023 bp in length
*    53017    53116: gap of unknown length
*    53117    54906: contig of 1790 bp in length
*    54907    55006: gap of unknown length
*    55007    56794: contig of 1788 bp in length
*    56795    56894: gap of unknown length
*    56895    58683: contig of 1789 bp in length
*    58684    58783: gap of unknown length
*    58784    60103: contig of 1320 bp in length
*    60104    60203: gap of unknown length
*    60204    61477: contig of 1274 bp in length
*    61478    61577: gap of unknown length
*    61578    62977: contig of 1400 bp in length
*    62978    63077: gap of unknown length
*    63078    64770: contig of 1693 bp in length
*    64771    64870: gap of unknown length
*    64871    66393: contig of 1523 bp in length
*    66394    66493: gap of unknown length
*    66494    67700: contig of 1207 bp in length
*    67701    67800: gap of unknown length
*    67801    68946: contig of 1146 bp in length
*    68947    69046: gap of unknown length
*    69047    70757: contig of 1711 bp in length
*    70758    70857: gap of unknown length
*    70858    72056: contig of 1199 bp in length
*    72057    72156: gap of unknown length
*    72157    73498: contig of 1342 bp in length
*    73499    73598: gap of unknown length
*    73599    75005: contig of 1407 bp in length
*    75006    75105: gap of unknown length
*    75106    76967: contig of 1862 bp in length
*    76968    77067: gap of unknown length
*    77068    78834: contig of 1767 bp in length
*    78835    78934: gap of unknown length
*    78935    80786: contig of 1852 bp in length
*    80787    80886: gap of unknown length
*    80887    82169: contig of 1283 bp in length
*    82170    82269: gap of unknown length
*    82270    83809: contig of 1540 bp in length
*    83810    83909: gap of unknown length
*    83910    85248: contig of 1339 bp in length
*    85249    85348: gap of unknown length

  Query Match             17.1%;  Score 84.2;  DB 2;  Length 163381;
  Best Local Similarity   52.3%;  Pred. No. 5.6e-06;
  Matches  172;  Conservative    0;  Mismatches  156;  Indels    1;  Gaps    1;

Qy      19 CCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTC 78
           || | ||  | |      |              || || |   | || ||||    | | |
Db   83473 CCCCTCCCNCCTNNNCNNCNNNNNNNNNGNNCCCCCCCCCNNNCCCCCCCCCCCCCCCCCC 83532

Qy      79 CTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCC 138
           | |  || || || ||| | |  || || | |||  ||||  || | || || | |   | |
Db   83533 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCNCCGC 83592

Qy     139 TAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTT 198
              |||  | | | |  || |  ||   |||| ||| || ||| | ||  ||||| |
Db   83593 CCCNCCCNCCCCCCCCCNCCCCCGCCCGGGGCGCGGCCGCGGCCGCGGCGCGCGCGGCCCG 83652

Qy     199 GAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGC 258
             | | | | ||||||| |  |  | |  |||   |||   ||  ||| |  | |||||
Db   83653 CCGGGGGCGGGGGCGCCGGGGGGCGGGGCCGGGGCGGGGGGCG-GCGCGGGGGCGGCGGG 83711

Qy     259 CGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCA 318
           || || | | |  ||||| | ||| |     | | ||| |||||  |  |  ||   |||
Db   83712 CGCCCCGCGCGGCGCGGCGGGGGGCGGCGGCGGGGCGGGGCGGCGGNGGCCGGGCCCGCG 83771

Qy     319 ACGGGCGGCGGACGCTGCTGGGGAGGTGC 347
```

```
           |||  |   | ||  || || |  |  |
Db  83772 NCGGCGGCNCGGCGGGGCGGGCGCCGCNC 83800


RESULT 4
AC107145
LOCUS       AC107145               177168 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-106E18, *** SEQUENCING IN PROGRESS
            ***, 67 unordered pieces.
ACCESSION   AC107145
VERSION     AC107145.3  GI:21737159
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 177168)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177168)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 177168)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18701851.
```

```
-------------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: GOQJ
    Center clone name: CH230-106E18
-------------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 120304 bases at least Q40
    Consensus quality: 126448 bases at least Q30
    Consensus quality: 131535 bases at least Q20
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1020: contig of 1020 bp in length
*     1021     1120: gap of unknown length
*     1121     2478: contig of 1358 bp in length
*     2479     2578: gap of unknown length
*     2579     3589: contig of 1011 bp in length
*     3590     3689: gap of unknown length
*     3690     4826: contig of 1137 bp in length
*     4827     4926: gap of unknown length
*     4927     6143: contig of 1217 bp in length
*     6144     6243: gap of unknown length
*     6244     7578: contig of 1335 bp in length
*     7579     7678: gap of unknown length
*     7679     8692: contig of 1014 bp in length
*     8693     8792: gap of unknown length
*     8793     9896: contig of 1104 bp in length
*     9897     9996: gap of unknown length
*     9997    11336: contig of 1340 bp in length
*    11337    11436: gap of unknown length
*    11437    13217: contig of 1781 bp in length
*    13218    13317: gap of unknown length
*    13318    15162: contig of 1845 bp in length
*    15163    15262: gap of unknown length
*    15263    16473: contig of 1211 bp in length
*    16474    16573: gap of unknown length
*    16574    17582: contig of 1009 bp in length
*    17583    17682: gap of unknown length
*    17683    18900: contig of 1218 bp in length
*    18901    19000: gap of unknown length
*    19001    20097: contig of 1097 bp in length
*    20098    20197: gap of unknown length
*    20198    21776: contig of 1579 bp in length
*    21777    21876: gap of unknown length
*    21877    23292: contig of 1416 bp in length
*    23293    23392: gap of unknown length
*    23393    24939: contig of 1547 b  in length
*    24940    25039: gap of unknown l ngth
*    25040    27311: contig of 2272 bp in length
*    27312    27411: gap of unknown length
*    27412    28968: contig of 1557 bp in length
*    28969    29068: gap of unknown length
*    29069    30742: contig of 1674 bp in length
*    30743    30842: gap of unknown length
*    30843    32409: contig of 1567 bp in length
*    32410    32509: gap of unknown length
*    32510    34377: contig of 1868 bp in length
*    34378    34477: gap of unknown length
*    34478    35582: contig of 1105 bp in length
*    35583    35682: gap of unknown length
*    35683    37613: contig of 1931 bp in length
*    37614    37713: gap of unknown length
*    37714    39152: contig of 1439 bp in length
*    39153    39252: gap of unknown length
*    39253    40589: contig of 1337 bp in length
*    40590    40689: gap of unknown length
*    40690    42099: contig of 1410 bp in length
*    42100    42199: gap of unknown length
*    42200    43365: contig of 1166 bp in length
*    43366    43465: gap of unknown length
*    43466    44469: contig of 1004 bp in length
*    44470    44569: gap of unknown length
*    44570    45636: contig of 1067 bp in length
*    45637    45736: gap of unknown length
*    45737    47826: contig of 2090 bp in length
*    47827    47926: gap of unknown length
*    47927    49153: contig of 1227 bp in length
*    49154    49253: gap of unknown length
*    49254    51098: contig of 1845 bp in length
*    51099    51198: gap of unknown length
*    51199    53492: contig of 2294 bp in length
*    53493    53592: gap of unknown length
*    53593    56557: contig of 2965 bp in length
*    56558    56657: gap of unknown length
*    56658    58011: contig of 1354 bp in length
*    58012    58111: gap of unknown length
*    58112    60168: contig of 2057 bp in length
*    60169    60268: gap of unknown length
*    60269    62494: contig of 2226 bp in length
*    62495    62594: gap of unknown length
*    62595    64041: contig of 1447 bp in length
*    64042    64141: gap of unknown length
*    64142    66152: contig of 2011 bp in length
*    66153    66252: gap of unknown length
*    66253    68551: contig of 2299 bp in length
*    68552    68651: gap of unknown length
*    68652    7?360: contig of 1709 bp in length
*    70361    7 160: gap of unknown leng h
*    7( 461   /2984: contig of 2524 bp in length
*    72985    73084: gap of unknown length
*    73085    76398: contig of 3314 bp in length
*    76399    76498: gap of unknown length
*    76499    79360: contig of 2862 bp in length
*    79361    79460: gap of unknown length
*    79461    81701: contig of 2241 bp in length
*    81702    81801: gap of unknown length
*    81802    85349: con(  f 3548 bp in length
*    85350    85449: gap of unknown length
*    85450    88601: contig of 3152 bp in length
*    88602    88701: gap of unknown length
*    88702    92201: contig of 3500 bp in length
*    92202    92301: gap of unknown length
*    92302    96407: contig of 4106 bp in length
*    96408    96507: gap of unknown length
*    96508   101699: contig of 5192 bp in length
*   101700   101799: gap of unknown length
*   101800   105571: contig of 3772 bp in length
*   105572   105671: gap of unknown length
*   105672   108750: contig of 3079 bp in length

  Query Match             15.9%;  Score 78;  DB 2;  Length 177168;
  Best Local Similarity   47.1%;  Pred. No. 7.6e-05;
  Matches  156;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy      19 CCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTC 78
           || |||           ||   | ||       | || ||  | |  ||||    | | |
Db   81318 CCCCCCNNNNCCCCNNCCCCNCCCCCNNNNCCNCCCCCCCCCCCCCNNCCCCCNCCCCCNC 81377

Qy      79 CTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCC 138
           | |  || || || ||| |    || || | |||  || |  || | || || |    ||
Db   81378 CCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCNCCNCCCCCCCCCCCCCCCCCGCC 81437
```

```
Qy     139 TAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTT 198
           |||     | || || ||    |  | || |  | |||  |    | ||||
Db   81438 CNNNNCCCNNNNCCCCCCCCCCCCNNNNCNCCCCCCCNNCCCCCNCGGGCCCGGCGGCGCN 81497

Qy     199 GAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGC 258
           |   ||| | |||| |   ||     |   ||   ||||| ||   | ||| ||||| ||
Db   81498 GGCGGCGCGNGGGCCCGGCCCCGGCGGGGGGGGGGGGGCCCCGNNGGGCGCGGGGGCCGC 81557

Qy     259 CGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCA 318
           |||| ||   |    ||  | |||||     ||| |   |  || || || |     |
Db   81558 CGGCGGGCNGGNGNGGGGCGGGGGGGGNCCCGCGCCCCGGGCGCGGGGGGGCCCCGGGG 81617

Qy     319 ACGGGCGGCGGACGCTGCTGGGGAGGTGCTC 349
           || |    ||     |   ||| |  || |
Db   81618 CCGCGGCCGGGCNCNCGGGCGGGGGCNGCGC 81648


RESULT 5
AC111697
LOCUS       AC111697               138467 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-20C20, *** SEQUENCING IN PROGRESS
            ***, 59 unordered pieces.
ACCESSION   AC111697
VERSION     AC111697.2  GI:21736261
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 138467)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138467)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 138467)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18701464.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMWR
                Center clone name: CH230-20C20
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 97272 bases at least Q40
                Consensus quality: 104545 bases at least Q30
                Consensus quality: 110146 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 59 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1060: contig of 1060 bp in length
            *     1061     1160: gap of unknown length
            *     1161     2228: contig of 1068 bp in length
            *     2229     2328: gap of unknown length
            *     2329     3520: contig of 1192 bp in length
            *     3521     3620: gap of unknown length
            *     3621     4797: contig of 1177 bp in length
            *     4798     4897: gap of unknown length
            *     4898     6624: contig of 1727 bp in length
            *     6625     6724: gap of unknown length
            *     6725     8301: contig of 1577 bp in length
            *     8302     8401: gap of unknown length
            *     8402     9813: contig of 1412 bp in length
            *     9814     9913: gap of unknown length
            *     9914    11128: contig of 1215 bp in length
            *    11129    11228: gap of unknown length
            *    11229    12658: contig of 1430 bp in length
            *    12659    12758: gap of unknown length
            *    12759    15360: contig of 2602 bp in length
            *    15361    15460: gap of unknown length
            *    15461    17025: contig of 1565 bp in length
            *    17026    17125: gap of unknown length
            *    17126    18259: contig of 1134 bp in length
            *    18260    18359: gap of unknown length
            *    18360    19979: contig of 1620 bp in length
            *    19980    20079: gap of unknown length
            *    20080    21208: contig of 1129 bp in length
            *    21209    21308: gap of unknown length
            *    21309    22587: contig of 1279 bp in length
            *    22588    22687: gap of unknown length
            *    22688    24387: contig of 1700 bp in length
            *    24388    24487: gap of unknown length
            *    24488    26131: contig of 1644 bp in length
            *    26132    26231: gap of unknown length
```

```
*    26232    27699: contig of 1468 bp in length
*    27700    27799: gap of unknown length
*    27800    29145: contig of 1346 bp in length
*    29146    29245: gap of unknown length
*    29246    30337: contig of 1092 bp in length
*    30338    30437: gap of unknown length
*    30438    31510: contig of 1073 bp in length
*    31511    31610: gap of unknown length
*    31611    33147: contig of 1537 bp in length
*    33148    33247: gap of unknown length
*    33248    34743: contig of 1496 bp in length
*    34744    34843: gap of unknown length
*    34844    36324: contig of 1481 bp in length
*    36325    36424: gap of unknown length
*    36425    37988: contig of 1564 bp in length
*    37989    38088: gap of unknown length
*    38089    40522: contig of 2434 bp in length
*    40523    40622: gap of unknown length
*    40623    41876: contig of 1254 bp in length
*    41877    41976: gap of unknown length
*    41977    43465: contig of 1489 bp in length
*    43466    43565: gap of unknown length
*    43566    45265: contig of 1700 bp in length
*    45266    45365: gap of unknown length
*    45366    47618: contig of 2253 bp in length
*    47619    47718: gap of unknown length
*    47719    49624: contig of 1906 bp in length
*    49625    49724: gap of unknown length
*    49725    51701: contig of 1977 bp in length
*    51702    51801: gap of unknown length
*    51802    54174: contig of 2373 bp in length
*    54175    54274: gap of unknown length
*    54275    56347: contig of 2073 bp in length
*    56348    56447: gap of unknown length
*    56448    57737: contig of 1290 bp in length
*    57738    57837: gap of unknown length
*    57838    60324: contig of 2487 bp in length
*    60325    60424: gap of unknown length
*    60425    61731: contig of 1307 bp in length
*    61732    61831: gap of unknown length
*    61832    63545: contig of 1714 bp in length
*    63546    63645: gap of unknown length
*    63646    65262: contig of 1617 bp in length
*    65263    65362: gap of unknown length
*    65363    67811: contig of 2449 bp in length
*    67812    67911: gap of unknown length
*    67912    70170: contig of 2259 bp in length
*    70171    70270: gap of unknown length
*    70271    72210: contig of 1940 bp in length
*    72211    72310: gap of unknown length
*    72311    73697: contig of 1387 bp in length
*    73698    73797: gap of unknown length
*    73798    75398: contig of 1601 bp in length
*    75399    75498: gap of unknown length
*    75499    78702: contig of 3204 bp in length
*    78703    78802: gap of unknown length
*    78803    81340: contig of 2538 bp in length
*    81341    81440: gap of unknown length
*    81441    84907: contig of 3467 bp in length
*    84908    85007: gap of unknown length
*    85008    88196: contig of 3189 bp in length
*    88197    88296: gap of unknown length
*    88297    91902: contig of 3606 bp in length
*    91903    92002: gap of unknown length
*    92003    95117: contig of 3115 bp in length
*    95118    95217: gap of unknown length
*    95218    98243: contig of 3026 bp in length
*    98244    98343: gap of unknown length
*    98344   102722: contig of 4379 bp in length
*   102723   102822: gap of unknown length
*   102823   105855: contig of 3033 bp in length
*   105856   105955: gap of unknown length
*   105956   110171: contig of 4216 bp in length
```

```
  Query Match             15.6%;  Score 76.8;  DB 2;  Length 138467;
  Best Local Similarity   52.4%;  Pred. No. 0.00013;
  Matches  176;  Conservative    0;  Mismatches  159;  Indels    1;  Gaps    1;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | || ||||          ||     |||  |  || || ||  | || ||||    | |
Db    2848 CNCCCCCCCCCNCCCCCNCCCNNNNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 2907

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
             |||  || || || ||| | |  || || | |||  ||||  |  | || || |   |
Db    2908 CTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCC 2967

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | | || || ||  | || | |||||    ||| |     || |  |
Db    2968 CCCCGCCCCCCCCCCCCCCCCCCCCCCGCCCCNCCCCGCCCNCGCCGGGGGCGGCGGGCGC 3027

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
             ||||   ||||| | |  |  | |   ||   | | | ||  |||||     ||
Db    3028 CGCGCCGCGCCGGGGCGGCCGGGGCGGGGCGGGGCCGGCGGCCCGCCCGCCGGCCCCGCC 3087

Qy     257 GCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCG 316
           | || |    | || || || | ||||| |  || | | |  | || || || || || |
Db    3088 GGCGCCGCCCGGGCCGC-GCGGGGGGGGCGGCGGGGGCCGCCCCGCGGGCGGGGGGGGGG 3146

Qy     317 CAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANG 352
           ||||||||  |  |    || || || |  |
Db    3147 GGCGCGGCGGCGGCGGGGGGGCCGGGGGGGCGCGCG 3182


RESULT 6
AC062001
LOCUS       AC062001               82815 bp    DNA     linear   HTG 21-APR-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-124H2 map 4, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC062001
VERSION     AC062001.1  GI:7630740
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 82815)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-124H2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 82815)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
```

```
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L9871
    Center clone name: 124_H_2
--------------.
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      975: contig of 975 bp in length
*      976 1075: gap of      100 bp
*     1076     2100: contig of 1025 bp in length
*     2101 2200: gap of      100 bp
*     2201     3185: contig of 985 bp in length
*     3186 3285: gap of      100 bp
*     3286     4151: contig of 866 bp in length
*     4152 4251: gap of      100 bp
*     4252     5206: contig of 955 bp in length
*     5207 5306: gap of      100 bp
*     5307     6287: contig of 981 bp in length
*     6288 6387: gap of      100 bp
*     6388     7332: contig of 945 bp in length
*     7333 7432: gap of      100 bp
*     7433     8397: contig of 965 bp in length
*     8398 8497: gap of      100 bp
*     8498     9457: contig of 960 bp in length
*     9458 9557: gap of      100 bp
*     9558    10489: contig of 932 bp in length
*    10490 10589: gap of      100 bp
*    10590    11602: contig of 1013 bp in length
*    11603 11702: gap of      100 bp
*    11703    12694: contig of 992 bp in length
*    12695 12794: gap of      100 bp
*    12795    13798: contig of 1004 bp in length
*    13799 13898: gap of      100 bp
*    13899    14866: contig of 968 bp in length
*    14867 14966: gap of      100 bp
*    14967    15930: contig of 964 bp in length
*    15931 16030: gap of      100 bp
*    16031    16981: contig of 951 bp in length
*    16982 17081: gap of      100 bp
*    17082    18043: contig of 962 bp in length
*    18044 18143: gap of      100 bp
*    18144    19119: contig of 976 bp in length
*    19120 19219: gap of      100 bp
*    19220    20173: contig of 954 bp in length
*    20174 20273: gap of      100 bp
*    20274    21232: contig of 959 bp in length
*    21233 21332: gap of      100 bp
*    21333    22227: contig of 895 bp in length
*    22228 22327: gap of      100 bp
*    22328    23303: contig of 976 bp in length
*    23304 23403: gap of      100 bp
*    23404    24393: contig of 990 bp in length
*    24394 24493: gap of      100 bp
*    24494    25495: contig of 1002 bp in length
*    25496 25595: gap of      100 bp
*    25596    26554: contig of 959 bp in length
*    26555 26654: gap of      100 bp
*    26655    27611: contig of 957 bp in length
*    27612 27711: gap of      100 bp
*    27712    28689: contig of 978 bp in length
*    28690 28789: gap of      100 bp
*    28790    29750: contig of 961 bp in length
*    29751 29850: gap of      100 bp
*    29851    30790: contig of 940 bp in length
*    30791 30890: gap of      100 bp
*    30891    31881: contig of 991 bp in length
*    31882 31981: gap of      100 bp
*    31982    32962: contig of 981 bp in length
*    32963 33062: gap of      100 bp
*    33063    34046: contig of 984 bp in length
*    34047 34146: gap of      100 bp
*    34147    35108: contig of 962 bp in length
*    35109 35208: gap of      100 bp
*    35209    36136: contig of 928 bp in length
*    36137 36236: gap of      100 bp
*    36237    37202: contig of 966 bp in length
*    37203 37302: gap of      100 bp
*    37303    38272: contig of 970 bp in length
*    38273 38372: gap of      100 bp
*    38373    39319: contig of 947 bp in length
*    39320 39419: gap of      100 bp
*    39420    40376: contig of 957 bp in length
*    40377 40476: gap of      100 bp
*    40477    41464: contig of 988 bp in length
*    41465 41564: gap of      100 bp
*    41565    42492: contig of 928 bp in length
*    42493 42592: gap of      100 bp
*    42593    43578: contig of 986 bp in length
*    43579 43678: gap of      100 bp
*    43679    44686: contig of 1008 bp in length
*    44687 44786: gap of      100 bp
*    44787    45745: contig of 959 bp in length
*    45746 45845: gap of      100 bp
*    45846    46827: contig of 982 bp in length
*    46828 46927: gap of      100 bp
*    46928    47910: contig of 983 bp in length
*    47911 48010: gap of      100 bp
*    48011    48969: contig of 959 bp in length
*    48970 49069: gap of      100 bp
*    49070    50028: contig of 959 bp in length
*    50029 50128: gap of      100 bp
*    50129    51053: contig of 925 bp in length
*    51054 51153: gap of      100 bp
*    51154    52111: contig of 958 bp in length
*    52112 52211: gap of      100 bp
*    52212    53189: contig of 978 bp in length
*    53190 53289: gap of      100 bp
*    53290    54271: contig of 982 bp in length
*    54272 54371: gap of      100 bp
*    54372    55336: contig of 965 bp in length
*    55337 55436: gap of      100 bp
*    55437    56366: contig of 930 bp in length
*    56367 56466: gap of      100 bp
*    56467    57415: contig of 949 bp in length
*    57416 57515: gap of      100 bp
*    57516    58501: contig of 986 bp in length
*    58502 58601: gap of      100 bp
*    58602    59567: contig of 966 bp in length
*    59568 59667: gap of      100 bp
*    59668    60585: contig of 918 bp in length
*    60586 60685: gap of      100 bp
*    60686    61646: contig of 961 bp in length
*    61647 61746: gap of      100 bp
*    61747    62739: contig of 993 bp in length
*    62740 62839: gap of      100 bp
*    62840    63824: contig of 985 bp in length
*    63825 63924: gap of      100 bp
*    63925    64903: contig of 979 bp in length
*    64904 65003: gap of      100 bp
*    65004    65963: contig of 960 bp in length
*    65964 66063: gap of      100 bp
*    66064    67010: contig of 947 bp in length
```

```
            *    67011 67110: gap of      100 bp
            *    67111    68077: contig of 967 bp in length
            *    68078 68177: gap of      100 bp
            *    68178    69147: contig of 970 bp in length
            *    69148 69247: gap of      100 bp
            *    69248    70258: contig of 1011 bp in length
            *    70259 70358: gap of      100 bp
            *    70359    71286: contig of 928 bp in length
            *    71287 71386: gap of      100 bp
            *    71387    72279: contig of 893 bp in length
            *    72280 72379: gap of      100 bp
            *    72380    73317: contig of 938 bp in length
            *    73318 73417: gap of      100 bp

  Query Match             15.1%;  Score 74.4;  DB 2;  Length 82815;
  Best Local Similarity   48.9%;  Pred. No. 0.00036;
  Matches  159;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | || ||||  |       ||   | | |  |  || || ||  | || | ||    | |
Db   66610 CCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNCCCCCCCCCCCCCCCCNCCNCCCCCC 66669

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
            ||  |  || || || ||| | |   | || | |||  ||||  || | |  || |
Db   66670 CCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCGCN 66729

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | | || || || ||    | ||    | || |   |   ||||
Db   66730 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNNCCCCCNNNCCCCGGCCCCNNNCGCGNNN 66789

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCC  TGGGGCG 256
           | | ||| | || |     | |     |    |     || |  ||  |||| |    |||
Db   66790 CGGCGGGCGCGGGGCCCGGGCCGGGCNGGGGNGCGCCGGCCGCCGGGCGCCCCGCCCGCG 66849

Qy     257 GCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCG 316
           |  ||| || | ||    ||   |||||     ||  |||  |||  || |  || || |
Db   66850 GGGGGCGGGGGGGCGCGCGCCCCGGGGGNNGCCGCCGCGGCTCGGGCGGGGCGGGGGGGG 66909

Qy     317 CAACGGGCGGCGGACGCTGCTGGGG 341
           |     |||  |  |   || || |
Db   66910 CCGGCNGCGCGGCCCCGCGCCGGCG 66934


RESULT 7
AC121439
LOCUS       AC121439              158897 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-247E4, *** SEQUENCING IN PROGRESS
            ***, 56 unordered pieces.
ACCESSION   AC121439
VERSION     AC121439.2  GI:21321741
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 158897)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158897)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 158897)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 4, 2002 this sequence version replaced gi:20976389.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: GXXP
                Center clone name: CH230-247E4
            ------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 125365 bases at least Q40
                Consensus quality: 129911 bases at least Q30
                Consensus quality: 132859 bases at least Q20
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 56 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1137: contig of 1137 bp in length
            *     1138     1237: gap of unknown length
            *     1238     2390: contig of 1153 bp in length
            *     2391     2490: gap of unknown length
            *     2491     3583: contig of 1093 bp in length
            *     3584     3683: gap of unknown length
            *     3684     4820: contig of 1137 bp in length
            *     4821     4920: gap of unknown length
```

```
*     4921     6335: contig of 1415 bp in length
*     6336     6435: gap of unknown length
*     6436     7534: contig of 1099 bp in length
*     7535     7634: gap of unknown length
*     7635     9844: contig of 2210 bp in length
*     9845     9944: gap of unknown length
*     9945    11234: contig of 1290 bp in length
*    11235    11334: gap of unknown length
*    11335    12344: contig of 1010 bp in length
*    12345    12444: gap of unknown length
*    12445    13678: contig of 1234 bp in length
*    13679    13778: gap of unknown length
*    13779    15477: contig of 1699 bp in length
*    15478    15577: gap of unknown length
*    15578    16737: contig of 1160 bp in length
*    16738    16837: gap of unknown length
*    16838    19340: contig of 2503 bp in length
*    19341    19440: gap of unknown length
*    19441    20568: contig of 1128 bp in length
*    20569    20668: gap of unknown length
*    20669    22352: contig of 1684 bp in length
*    22353    22452: gap of unknown length
*    22453    23795: contig of 1343 bp in length
*    23796    23895: gap of unknown length
*    23896    25953: contig of 2058 bp in length
*    25954    26053: gap of unknown length
*    26054    28136: contig of 2083 bp in length
*    28137    28236: gap of unknown length
*    28237    30151: contig of 1915 bp in length
*    30152    30251: gap of unknown length
*    30252    31447: contig of 1196 bp in length
*    31448    31547: gap of unknown length
*    31548    32949: contig of 1402 bp in length
*    32950    33049: gap of unknown length
*    33050    34438: contig of 1389 bp in length
*    34439    34538: gap of unknown length
*    34539    36281: contig of 1743 bp in length
*    36282    36381: gap of unknown length
*    36382    37951: contig of 1570 bp in length
*    37952    38051: gap of unknown length
*    38052    39294: contig of 1243 bp in length
*    39295    39394: gap of unknown length
*    39395    40924: contig of 1530 bp in length
*    40925    41024: gap of unknown length
*    41025    42151: contig of 1127 bp in length
*    42152    42251: gap of unknown length
*    42252    43924: contig of 1673 bp in length
*    43925    44024: gap of unknown length
*    44025    46633: contig of 2609 bp in length
*    46634    46733: gap of unknown length
*    46734    48320: contig of 1587 bp in length
*    48321    48420: gap of unknown length
*    48421    50491: contig of 2071 bp in length
*    50492    50591: gap of unknown length
*    50592    53057: contig of 2466 bp in length
*    53058    53157: gap of unknown length
*    53158    56233: contig of 3076 bp in length
*    56234    56333: gap of unknown length
*    56334    59001: contig of 2668 bp in length
*    59002    59101: gap of unknown length
*    59102    61813: contig of 2712 bp in length
*    61814    61913: gap of unknown length
*    61914    66256: contig of 4343 bp in length
*    66257    66356: gap of unknown length
*    66357    68713: contig of 2357 bp in length
*    68714    68813: gap of unknown length
*    68814    70772: contig of 1959 bp in length
*    70773    70872: gap of unknown length
*    70873    74228: contig of 3356 bp in length
*    74229    74328: gap of unknown length
*    74329    77316: contig of 2988 bp in length
*    77317    77416: gap of unknown length
*    77417    81174: contig of 3758 bp in length
*    81175    81274: gap of unknown length
*    81275    85927: contig of 4653 bp in length
*    85928    86027: gap of unknown length
*    86028    89056: contig of 3029 bp in length
*    89057    89156: gap of unknown length
*    89157    93557: contig of 4401 bp in length
*    93558    93657: gap of unknown length
*    93658    96490: contig of 2833 bp in length
*    96491    96590: gap of unknown length
*    96591   101020: contig of 4430 bp in length
*   101021   101120: gap of unknown length
*   101121   104653: contig of 3533 bp in length
*   104654   104753: gap of unknown length
*   104754   108393: contig of 3640 bp in length
*   108394   108493: gap of unknown length
*   108494   112698: contig of 4205 bp in length
*   112699   112798: gap of unknown length
*   112799   118578: contig of 5780 bp in length
*   118579   118678: gap of unknown length
*   118679   124462: contig of 5784 bp in length
*   124463   124562: gap of unknown length
*   124563   129728: contig of 5166 bp in length
*   129729   129828: gap of unknown length
*   129829   135194: contig of 5366 bp in length
*   135195   135294: gap of unknown length
*   135295   143870: contig of 8576 bp in length

  Query Match             14.9%;  Score 73.2;  DB 2;  Length 158897;
  Best Local Similarity   49.9%;  Pred. No. 0.00058;
  Matches  183;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy      10 CGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCC 69
           | || | |||  | ||  | ||    |||    | |  |  || || ||  ||| |||||
Db   35832 CTTCTTCCTCTCCTCCTTCCTTCCTTCCTCCTCCTCTCCCTCCCCCCCCCCCCTCTTCCCC 35891

Qy      70 AATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTC 129
             ||  | || |||||||| ||| |||   | | || | || |||   | |  ||||
Db   35892 CCCTCTTCCTTCCTCCTCCTCCCTCTCTCTCCCTTCTCTCTCCCCTTTCTTTCCCTC 35951

Qy     130 CGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTC 189
             |    |||   |||| |||||  || | || || |  |  | ||||  |  | |   |
Db   35952 TTCTTCTCCTCCTCCCCTCTCTCCTCTGCCCCTCCCCTCCTTCCCCTCCTCTTTCTTCCC 36011

Qy     190 CGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGC 249
             || |  ||    |  |  |           | |    |  ||| | || |
Db   36012 TTCGCTGCTGGCGCCCCCTTCTCTGTCCGGTCGGCGGCGCGGCGGGGGCGCGGTGTTGTG 36071

Qy     250 TGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGA 309
             | |  |  || |||  | |  | ||  | ||| | |  || |  || | ||| ||||
Db   36072 GGTGCGGCTCGTCCGCTGCGGCGGGGGGGGGGGGGCGTGCGGGGGGGGGGGGGGCGGGTGTG 36131

Qy     310 GGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCG 369
           || || |    |||  | ||     | ||||| |  | ||  |   |      | | |||
Db   36132 GGGGGGGGTTGGGGGTGAGGGGCGGGGTGGGGGGTGGGTCGCGCGCTGTGGTGGGCGGCG 36191

Qy     370 GGTTGAG 376
           ||| | |
Db   36192 GGTGGCG 36198


RESULT 8
AC116109
LOCUS       AC116109               68330 bp    DNA     linear   HTG 25-MAR-2002
DEFINITION  Mus musculus clone RP23-467G14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC116109
VERSION     AC116109.1  GI:19703220
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 68330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
  TITLE     Mus musculus, clone RP23-467G14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68330)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trujilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L22810
                Center clone name: 467_G_14
            --------------.
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*        1      647: contig of 647 bp in length
*      648 747: gap of      100 bp
*      748     1422: contig of 675 bp in length
*     1423 1522: gap of      100 bp
*     1523     2229: contig of 707 bp in length
*     2230 2329: gap of      100 bp
*     2330     3001: contig of 672 bp in length
*     3002 3101: gap of      100 bp
*     3102     3783: contig of 682 bp in length
*     3784 3883: gap of      100 bp
*     3884     4586: contig of 703 bp in length
*     4587 4686: gap of      100 bp
*     4687     5381: contig of 695 bp in length
*     5382 5481: gap of      100 bp
*     5482     6190: contig of 709 bp in length
*     6191 6290: gap of      100 bp
*     6291     7003: contig of 713 bp in length
*     7004 7103: gap of      100 bp
*     7104     7824: contig of 721 bp in length
*     7825 7924: gap of      100 bp
*     7925     8662: contig of 738 bp in length
*     8663 8762: gap of      100 bp
*     8763     9402: contig of 640 bp in length
*     9403 9502: gap of      100 bp
*     9503    10235: contig of 733 bp in length
*    10236 10335: gap of      100 bp
*    10336    11055: contig of 720 bp in length
*    11056 11155: gap of      100 bp
*    11156    11861: contig of 706 bp in length
*    11862 11961: gap of      100 bp
*    11962    12653: contig of 692 bp in length
*    12654 12753: gap of      100 bp
*    12754    13445: contig of 692 bp in length
*    13446 13545: gap of      100 bp
*    13546    14235: contig of 690 bp in length
*    14236 14335: gap of      100 bp
*    14336    15046: contig of 711 bp in length
*    15047 15146: gap of      100 bp
*    15147    15853: contig of 707 bp in length
*    15854 15953: gap of      100 bp
*    15954    16632: contig of 679 bp in length
*    16633 16732: gap of      100 bp
*    16733    17441: contig of 709 bp in length
*    17442 17541: gap of      100 bp
*    17542    18270: contig of 729 bp in length
*    18271 18370: gap of      100 bp
*    18371    19048: contig of 678 bp in length
*    19049 19148: gap of      100 bp
*    19149    19851: contig of 703 bp in length
*    19852 19951: gap of      100 bp
*    19952    20661: contig of 710 bp in length
*    20662 20761: gap of      100 bp
*    20762    21473: contig of 712 bp in length
*    21474 21573: gap of      100 bp
*    21574    22286: contig of 713 bp in length
*    22287 22386: gap of      100 bp
*    22387    23086: contig of 700 bp in length
*    23087 23186: gap of      100 bp
*    23187    23920: contig of 734 bp in length
*    23921 24020: gap of      100 bp
*    24021    24713: contig of 693 bp in length
*    24714 24813: gap of      100 bp
*    24814    25540: contig of 727 bp in length
*    25541 25640: gap of      100 bp
*    25641    26364: contig of 724 bp in length
*    26365 26464: gap of      100 bp
*    26465    27180: contig of 716 bp in length
*    27181 27280: gap of      100 bp
*    27281    27977: contig of 697 bp in length
*    27978 28077: gap of      100 bp
*    28078    28769: contig of 692 bp in length
*    28770 28869: gap of      100 bp
*    28870    29559: contig of 690 bp in length
*    29560 29659: gap of      100 bp
*    29660    30357: contig of 698 bp in length
*    30358 30457: gap of      100 bp
*    30458    31165: contig of 708 bp in length
*    31166 31265: gap of      100 bp
*    31266    31948: contig of 683 bp in length
*    31949 32048: gap of      100 bp
*    32049    32768: contig of 720 bp in length
*    32769 32868: gap of      100 bp
*    32869    33548: contig of 680 bp in length
*    33549 33648: gap of      100 bp
*    33649    34357: contig of 709 bp in length
*    34358 34457: gap of      100 bp
*    34458    35187: contig of 730 bp in length
*    35188 35287: gap of      100 bp
*    35288    35990: contig of 703 bp in length
*    35991 36090: gap of      100 bp
*    36091    36767: contig of 677 bp in length
*    36768 36867: gap of      100 bp
*    36868    37575: contig of 708 bp in length
*    37576 37675: gap of      100 bp
*    37676    38379: contig of 704 bp in length
```

```
*    38380 38479: gap of      100 bp
*    38480    39200: contig of 721 bp in length
*    39201 39300: gap of      100 bp
*    39301    40023: contig of 723 bp in length
*    40024 40123: gap of      100 bp
*    40124    40827: contig of 704 bp in length
*    40828 40927: gap of      100 bp
*    40928    41659: contig of 732 bp in length
*    41660 41759: gap of      100 bp
*    41760    42482: contig of 723 bp in length
*    42483 42582: gap of      100 bp
*    42583    43307: contig of 725 bp in length
*    43308 43407: gap of      100 bp
*    43408    44124: contig of 717 bp in length
*    44125 44224: gap of      100 bp
*    44225    44923: contig of 699 bp in length
*    44924 45023: gap of      100 bp
*    45024    45713: contig of 690 bp in length
*    45714 45813: gap of      100 bp
*    45814    46506: contig of 693 bp in length
*    46507 46606: gap of      100 bp
*    46607    47291: contig of 685 bp in length
*    47292 47391: gap of      100 bp
*    47392    48100: contig of 709 bp in length
*    48101 48200: gap of      100 bp
*    48201    48910: contig of 710 bp in length
*    48911 49010: gap of      100 bp
*    49011    49727: contig of 717 bp in length
*    49728 49827: gap of      100 bp
*    49828    50558: contig of 731 bp in length
*    50559 50658: gap of      100 bp
*    50659    51335: contig of 677 bp in length
*    51336 51435: gap of      100 bp
*    51436    52168: contig of 733 bp in length
*    52169 52268: gap of      100 bp
*    52269    52992: contig of 724 bp in length
*    52993 53092: gap of      100 bp
*    53093    53810: contig of 718 bp in length
*    53811 53910: gap of      100 bp
*    53911    54559: contig of 649 bp in length
*    54560 54659: gap of      100 bp
*    54660    55361: contig of 702 bp in length
*    55362 55461: gap of      100 bp

  Query Match             14.7%;  Score 72.2;  DB 2;  Length 68330;
  Best Local Similarity   52.3%;  Pred. No. 0.00093;
  Matches  137;  Conservative    0;  Mismatches  125;  Indels    0;  Gaps    0;

Qy      35 ACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACC 94
           ||       |      |  ||  |  | || ||||    |  | |  ||  || || ||
Db     304 ACNNNNNNNCNNCNCNCCCCNCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCC 363

Qy      95 CACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGC 154
           | | |  || || , |||  ||||  || | || || |   |||   ||||| | | | |
Db     364 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCC 423

Qy     155 CTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGC 214
           | || || || || | || || | | |     | | | ||    | |||||    |||||
Db     424 CCCCCCCCCCCCCCCCCCCCCCCCCGCCCCGGGGCGGGGGCGCGGGGCGCGCCCCGGCGC 483

Qy     215 CAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCG 274
              |||  | |   ||   |||    |   |  || |||||||||| ||| | |  |
Db     484 GGCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGCGGCCGGGCGGCGGGGGGGC 543

Qy     275 GCAGTGGGGGAGTTGGCGACGG 296
           |  |  |||| |    |  |||
Db     544 GGNGGNGGGGGGCCCCCCGCGG 565


RESULT 9
AC126076
LOCUS       AC126076              117082 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-143P6,  ** SEQUENCING IN PROGRESS
            ***, 63 unordered pieces.
ACCESSION   AC126076
VERSION     AC126076.2  GI:21699142
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 117082)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R. Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., ox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 117082)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 117082)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 5, 2002 this sequence version replaced gi:21686820.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GZTG
                Center clone name: CH230-143P6
```

```
-------------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 55139 bases at least Q40
    Consensus quality: 61060 bases at least Q30
    Consensus quality: 64630 bases at least Q20
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1055: contig of 1055 bp in length
*     1056     1155: gap of unknown length
*     1156     2165: contig of 1010 bp in length
*     2166     2265: gap of unknown length
*     2266     3314: contig of 1049 bp in length
*     3315     3414: gap of unknown length
*     3415     5018: contig of 1604 bp in length
*     5019     5118: gap of unknown length
*     5119     6243: contig of 1125 bp in length
*     6244     6343: gap of unknown length
*     6344     7662: contig of 1319 bp in length
*     7663     7762: gap of unknown length
*     7763     9164: contig of 1402 bp in length
*     9165     9264: gap of unknown length
*     9265    10352: contig of 1088 bp in length
*    10353    10452: gap of unknown length
*    10453    11637: contig of 1185 bp in length
*    11638    11737: gap of unknown length
*    11738    13085: contig of 1348 bp in length
*    13086    13185: gap of unknown length
*    13186    14617: contig of 1432 bp in length
*    14618    14717: gap of unknown length
*    14718    16407: contig of 1690 bp in length
*    16408    16507: gap of unknown length
*    16508    17676: contig of 1169 bp in length
*    17677    17776: gap of unknown length
*    17777    19207: contig of 1431 bp in length
*    19208    19307: gap of unknown length
*    19308    21017: contig of 1710 bp in length
*    21018    21117: gap of unknown length
*    21118    22172: contig of 1055 bp in length
*    22173    22272: gap of unknown length
*    22273    23366: contig of 1094 bp in length
*    23367    23466: gap of unknown length
*    23467    24549: contig of 1083 bp in length
*    24550    24649: gap of unknown length
*    24650    25665: contig of 1016 bp in length
*    25666    25765: gap of unknown length
*    25766    27421: contig of 1656 bp in length
*    27422    27521: gap of unknown length
*    27522    28564: contig of 1043 bp in length
*    28565    28664: gap of unknown length
*    28665    29755: contig of 1091 bp in length
*    29756    29855: gap of unknown length
*    29856    31657: contig of 1802 bp in length
*    31658    31757: gap of unknown length
*    31758    33595: contig of 1838 bp in length
*    33596    33695: gap of unknown length
*    33696    35198: contig of 1503 bp in length
*    35199    35298: gap of unknown length
*    35299    36466: contig of 1168 bp in length
*    36467    36566: gap of unknown length
*    36567    37640: contig of 1074 bp in length
*    37641    37740: gap of unknown length
*    37741    39003: contig of 1263 bp in length
*    39004    39103: gap of unknown length
*    39104    41305: contig of 2202 bp in length
*    41306    41405: gap of unknown length
*    41406    42620: contig of 1215 bp in length
*    42621    42720: gap of unknown length
*    42721    43940: contig of 1220 bp in length
*    43941    44040: gap of unknown length
*    44041    45996: contig of 1956 bp in length
*    45997    46096: gap of unknown length
*    46097    47529: contig of 1433 bp in length
*    47530    47629: gap of unknown length
*    47630    49451: contig of 1822 bp in length
*    49452    49551: gap of unknown length
*    49552    50801: contig of 1250 bp in length
*    50802    50901: gap of unknown length
*    50902    52881: contig of 1980 bp in length
*    52882    52981: gap of unknown length
*    52982    54499: contig of 1518 bp in length
*    54500    54599: gap of unknown length
*    54600    55934: contig of 1335 bp in length
*    55935    56034: gap of unknown length
*    56035    58138: contig of 2104 bp in length
*    58139    58238: gap of unknown length
*    58239    59732: contig of 1494 bp in length
*    59733    59832: gap of unknown length
*    59833    61285: contig of 1453 bp in length
*    61286    61385: gap of unknown length
*    61386    62880: contig of 1495 bp in length
*    62881    62980: gap of unknown length
*    62981    64930: contig of 1950 bp in length
*    64931    65030: gap of unknown length
*    65031    66998: contig of 1968 bp in length
*    66999    67098: gap of unknown length
*    67099    69768: contig of 2670 bp in length
*    69769    69868: gap of unknown length
*    69869    72122: contig of 2254 bp in length
*    72123    72222: gap of unknown length
*    72223    74144: contig of 1922 bp in length
*    74145    74244: gap of unknown length
*    74245    75957: contig of 1713 bp in length
*    75958    76057: gap of unknown length
*    76058    78577: contig of 2520 bp in length
*    78578    78677: gap of unknown length
*    78678    80255: contig of 1578 bp in length
*    80256    80355: gap of unknown length
*    80356    82593: contig of 2238 bp in length
*    82594    82693: gap of unknown length
*    82694    84685: contig of 1992 bp in length
*    84686    84785: gap of unknown length
*    84786    86448: contig of 1663 bp in length
*    86449    86548: gap of unknown length
*    86549    87884: contig of 1336 bp in length

  Query Match             14.7%;  Score 72.2;  DB 2;  Length 117082;
  Best Local Similarity   44.6%;  Pred. No. 0.00091;
  Matches  152;  Conservative    0;  Mismatches  189;  Indels    0;  Gaps    0;

Qy       9 GCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCC 68
           |                                                  ||  | ||  | || |||
Db   48749 GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCNCCCCCCCCCCCCCCC 48808

Qy      69 CAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCT 128
           |     | ||    || || |  ||| | |  || ||    ||  |||   || | ||
Db   48809 CNNCAAACCCCCNNNCCCCCCCNCCCCCCCCCCCCCCCCCCNNNCCGCCCCGCCCCCCCCCC 48868

Qy     129 CCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGT 188
           || |   |||   | ||| | | | ||  | | ||| ||||| || |      | | ||
Db   48869 CCCCCGCCCCCCCCNCCCCCCCCCCCCCNNCCCGGCCGCCGCGGCGGCGNCGNGGCGGCGG 48928

Qy     189 CCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCG 248
           | |  | |  | ||| | | ||| |   |  |  | |   ||   ||  |  |   | |
Db   48929 CGGGCGCCGCGCGCGGGCGGGGGGGGGGGGGGGCGGGGGGGGGGGGGCGCGGGGGGGGG 48988
```

```
Qy    249 CTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGG 308
          |||| ||  || ||| | |   |||    || || |  ||||  ||   ||     ||
Db  48989 GCGGGGGGGGGGGGGCGGGGGGGGCCGGGGCGGGCGGGGGGGGCGGGGGGCGGGGCCCGGG 49048

Qy    309 AGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTC 349
           ||  | |    ||| || ||  || || |||   | ||||
Db  49049 CGGGCGGGGGGGGGGGGGGGGGGGCGGCGGGGCGCGCGCTC 49089


RESULT 10
AC027353/c
LOCUS       AC027353               101509 bp    DNA     linear   HTG 26-SEP-2000
DEFINITION  Homo sapiens chromosome 16 clone RP11-167D21, WORKING DRAFT
            SEQUENCE, 68 unordered pieces.
ACCESSION   AC027353
VERSION     AC027353.4  GI:10305180
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 101509)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101509)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi:9090641.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-11_167D21
            -------
            Summary Statistics
            Consensus quality: 19476 bases at least Q40
            Consensus quality: 22258 bases at least Q30
            Consensus quality: 25343 bases at least Q20
            Estimated insert size: 26000; agarose-fp estimation
            Estimated insert size: 94809; sum-of-contigs estimation
            Quality coverage:  15.03 in Q20 bases; agarose-fp estimation
            Quality coverage:  4.12 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 68 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1470: contig of 1470 bp in length
            *     1471     1570: gap of unknown length
            *     1571     2728: contig of 1158 bp in length
            *     2729     2828: gap of unknown length
            *     2829     3955: contig of 1127 bp in length
            *      956     4055: gap of unknown length
            *     4056     5273: contig of 1218 bp in length
            *     5274     5373: gap of unknown length
            *     5374     6455: contig of 1082 bp in length
            *     6456     6555: gap of unknown length
            *     6556     7722: contig of 1167 bp in length
            *     7723     7822: gap of unknown length
            *     7823     8994: contig of 1172 bp in length
            *     8995     9094: gap of unknown length
            *     9095    10258: contig of 1164 bp in length
            *    10259    10358: gap of unknown length
            *    10359    11529: contig of 1171 bp in length
            *    11530    11629: gap of unknown length
            *    11630    12772: contig of 1143 bp in length
            *    12773    12872: gap of unknown length
            *    12873    13954: contig of 1082 bp in length
            *    13955    14054: gap of unknown length
            *    14055    15127: contig of 1073 bp in length
            *    15128    15227: gap of unknown length
            *    15228    16411: contig of 1184 bp in length
            *    16412    16511: gap of unknown length
            *    16512    17557: contig of 1046 bp in length
            *    17558    17657: gap of unknown length
            *    17658    18766: contig of 1109 bp in length
            *    18767    18866: gap of unknown length
            *    18867    20034: contig of 1168 bp in length
            *    20035    20134: gap of unknown length
            *    20135    21336: contig of 1202 bp in length
            *    21337    21436: gap of unknown length
            *    21437    22545: contig of 1109 bp in length
            *    22546    22645: gap of unknown length
            *    22646    23755: contig of 1110 bp in length
            *    23756    23855: gap of unknown length
            *    23856    24927: contig of 1072 bp in length
            *    24928    25027: gap of unknown length
            *    25028    26153: contig of 1126 bp in length
            *    26154    26253: gap of unknown length
            *    26254    27455: contig of 1202 bp in length
            *    27456    27555: gap of unknown length
            *    27556    28679: contig of 1124 bp in length
            *    28680    28779: gap of unknown length
            *    28780    29893: contig of 1114 bp in length
            *    29894    29993: gap of unknown length
            *    29994    31172: contig of 1179 bp in length
            *    31173    31272: gap of unknown length
            *    31273    32388: contig of 1116 bp in length
            *    32389    32488: gap of unknown length
            *    32489    33650: contig of 1162 bp in length
            *    33651    33750: gap of unknown length
            *    33751    34987: contig of 1237 bp in length
            *    34988    35087: gap of unknown length
            *    35088    36153: contig of 1066 bp in length
            *    36154    36253: gap of unknown length
            *    36254    37421: contig of 1168 bp in length
            *    37422    37521: gap of unknown length
            *    37522    38717: contig of 1196 bp in length
            *    38718    38817: gap of unknown length
            *    38818    39915: contig of 1098 bp in length
            *    39916    40015: gap of unknown length
            *    40016    41215: contig of 1200 bp in length
            *    41216    41315: gap of unknown length
            *    41316    42350: contig of 1035 bp in length
            *    42351    42450: gap of unknown length
            *    42451    43607: contig of 1157 bp in length
            *    43608    43707: gap of unknown length
            *    43708    44880: contig of 1173 bp in length
            *    44881    44980: gap of unknown length
            *    44981    46095: contig of 1115 bp in length
            *    46096    46195: gap of unknown length
            *    46196    47463: contig of 1268 bp in length
            *    47464    47563: gap of unknown length
            *    47564    48771: contig of 1208 bp in length
            *    48772    48871: gap of unknown length
            *    48872    50000: contig of 1129 bp in length
            *    50001    50100: gap of unknown length
            *    50101    51241: contig of 1141 bp in length
            *    51242    51341: gap of unknown length
            *    51342    52439: contig of 1098 bp in length
            *    52440    52539: gap of unknown length
            *    52540    53719: contig of 1180 bp in length
            *    53720    53819: gap of unknown length
            *    53820    54972: contig of 1153 bp in length
            *    54973    55072: gap of unknown length
            *    55073    56256: contig of 1184 bp in length
            *    56257    56356: gap of unknown length
```

```
*    56357    57447: contig of 1091 bp in length
*    57448    57547: gap of unknown length
*    57548    58710: contig of 1163 bp in length
*    58711    58810: gap of unknown length
*    58811    59999: contig of 1189 bp in length
*    60000    60099: gap of unknown length
*    60100    61343: contig of 1244 bp in length
*    61344    61443: gap of unknown length
*    61444    62568: contig of 1125 bp in length
*    62569    62668: gap of unknown length
*    62669    63904: contig of 1236 bp in length
*    63905    64004: gap of unknown length
*    64005    65045: contig of 1041 bp in length
*    65046    65145: gap of unknown length
*    65146    66360: contig of 1215 bp in length
*    66361    66460: gap of unknown length
*    66461    67705: contig of 1245 bp in length
*    67706    67805: gap of unknown length
*    67806    68914: contig of 1109 bp in length
*    68915    69014: gap of unknown length
*    69015    70199: contig of 1185 bp in length
*    70200    70299: gap of unknown length
*    70300    71355: contig of 1056 bp in length
*    71356    71455: gap of unknown length
*    71456    73215: contig of 1760 bp in length
*    73216    73315: gap of unknown length
*    73316    74456: contig of 1141 bp in length
*    74457    74556: gap of unknown length
*    74557    76312: contig of 1756 bp in length
*    76313    76412: gap of unknown length
*    76413    77484: contig of 1072 bp in length
*    77485    77584: gap of unknown length
*    77585    78683: contig of 1099 bp in length
*    78684    78783: gap of unknown length
*    78784    79858: contig of 1075 bp in length
*    79859    79958: gap of unknown length
*    79959    80987: contig of 1029 bp in length
*    80988    81087: gap of unknown length
*    81088    82188: contig of 1101 bp in length
*    82189    82288: gap of unknown length
*    82289    83320: contig of 1032 bp in length
*    83321    83420: gap of unknown length
*    83421    84521: contig of 1101 bp in length
*    84522    84621: gap of unknown length
*    84622   101509: contig of 16888 bp in length.
FEATURES             Location/Qualifiers
     source          1. .101509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-167D21"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    27989 a   7533 c  28794 g   7012 t  30181 others
ORIGIN

  Query Match           14.3%;  Score 70.4;  DB 2;  Length 101509;
  Best Local Similarity   43.3%;  Pred. No. 0.002;
  Matches  143;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | |  ||||  |       |    | |||  |  || || ||  | |   |||    | |
Db   31049 CNCNNCCCCCCCCGCCCCCCNNCNCCCCCCCCCCCCCCCCCCCCCCCNNNCCCCNCCCNC 30990

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
            ||  |  || || || |||   |  ||    | | |  ||||  || | || || |
Db   30989 CCCCCCCCCCCCCCCNCCCNNNCCCCCCNNCCCCNCNNCCCCCCCCCCCCCCCNCNNNN 30930

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | |  |       |  || |    || | |||    |  || |   |
Db   30929 CCCCCCCCCCCCCCCCCNNCNNNNNNNCNCCCCCCNNNCCCCNCCCCCCCCCCCCCNNCC 30870

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
                |       | |   |      |  |||   | |||  |   |      |||  |

Db   30869 CCCNNNCCCNCCNCNCNCNNNCNCCNCCGGCCGGCGCGNGCCGGGGCGGGGCGGCGGGGG 30810

Qy     257 GCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCG 316
           |||| |||| |  |  |  | |  |||| |  || |  |  | ||| |   | || || |
Db   30809 GCCGNCCGGGGGCCCCCCNCGGGCGGGGGGCGGGGGGGGCGGGGGCGGNCCGGGGGGGGG 30750

Qy     317 CAACGGGCGGCGGACGCTGCTGGGGAGGTG 346
               ||||         |  |   |  || |
Db   30749 NNCNGGGCCNNCCCNNCCCCNCNGCCGGGG 30720


RESULT 11
AC125906
LOCUS       AC125906               151358 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-77F8, *** SEQUENCING IN PROGRESS ***,
            63 unordered pieces.
ACCESSION   AC125906
VERSION     AC125906.1  GI:21671504
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 151358)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
    TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151358)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (02-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 151358)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: GVRX
              Center clone name: CH230-77F8
            -------------- Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 92473 bases at least Q40
              Consensus quality: 98495 bases at least Q30
              Consensus quality: 102505 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 63 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1032: contig of 1032 bp in length
            *     1033     1132: gap of unknown length
            *     1133     2134: contig of 1002 bp in length
            *     2135     2234: gap of unknown length
            *     2235     3255: contig of 1021 bp in length
            *     3256     3355: gap of unknown length
            *     3356     4398: contig of 1043 bp in length
            *     4399     4498: gap of unknown length
            *     4499     5857: contig of 1359 bp in length
            *     5858     5957: gap of unknown length
            *     5958     7206: contig of 1249 bp in length
            *     7207     7306: gap of unknown length
            *     7307     8379: contig of 1073 bp in length
            *     8380     8479: gap of unknown length
            *     8480     9982: contig of 1503 bp in length
            *     9983    10082: gap of unknown length
            *    10083    11373: contig of 1291 bp in length
            *    11374    11473: gap of unknown length
            *    11474    12776: contig of 1303 bp in length
            *    12777    12876: gap of unknown length
            *    12877    14025: contig of 1149 bp in length
            *    14026    14125: gap of unknown length
            *    14126    15578: contig of 1453 bp in length
            *    15579    15678: gap of unknown length
            *    15679    16957: contig of 1279 bp in length
            *    16958    17057: gap of unknown length
            *    17058    18241: contig of 1184 bp in length
            *    18242    18341: gap of unknown length
            *    18342    19373: contig of 1032 bp in length
            *    19374    19473: gap of unknown length
            *    19474    21050: contig of 1577 bp in length
            *    21051    21150: gap of unknown length
            *    21151    22862: contig of 1712 bp in length
            *    22863    22962: gap of unknown length
            *    22963    24563: contig of 1601 bp in length
            *    24564    24663: gap of unknown length
            *    24664    25683: contig of 1020 bp in length
            *    25684    25783: gap of unknown length
            *    25784    26831: contig of 1048 bp in length
            *    26832    26931: gap of unknown length
            *    26932    28243: contig of 1312 bp in length
            *    28244    28343: gap of unknown length
            *    28344    29758: contig of 1415 bp in length
            *    29759    29858: gap of unknown length
            *    29859    31174: contig of 1316 bp in length
            *    31175    31274: gap of unknown length
            *    31275    33433: contig of 2159 bp in length
            *    33434    33533: gap of unknown length
            *    33534    35127: contig of 1594 bp in length
            *    35128    35227: gap of unknown length
            *    35228    36393: contig of 1166 bp in length
            *    36394    36493: gap of unknown length
            *    36494    38614: contig of 2121 bp in length
            *    38615    38714: gap of unknown length
            *    38715    40665: contig of 1951 bp in length
            *    40666    40765: gap of unknown length
            *    40766    41885: contig of 1120 bp in length
            *    41886    41985: gap of unknown length
            *    41986    44086: contig of 2101 bp in length
            *    44087    44186: gap of unknown length
            *    44187    46007: contig of 1821 bp in length
            *    46008    46107: gap of unknown length
            *    46108    47865: contig of 1758 bp in length
            *    47866    47965: gap of unknown length
            *    47966    50021: contig of 2056 bp in length
            *    50022    50121: gap of unknown length
            *    50122    51983: contig of 1862 bp in length
            *    51984    52083: gap of unknown length
            *    52084    54300: contig of 2217 bp in length
            *    54301    54400: gap of unknown length
            *    54401    55910: contig of 1510 bp in length
            *    55911    56010: gap of unknown length
            *    56011    57542: contig of 1532 bp in length
            *    57543    57642: gap of unknown length
            *    57643    59295: contig of 1653 bp in length
            *    59296    59395: gap of unknown length
            *    59396    62393: contig of 2998 bp in length
            *    62394    62493: gap of unknown length
            *    62494    64229: contig of 1736 bp in length
            *    64230    64329: gap of unknown length
            *    64330    66089: contig of 1760 bp in length
            *    66090    66189: gap of unknown length
            *    66190    68909: contig of 2720 bp in length
            *    68910    69009: gap of unknown length
            *    69010    71491: contig of 2482 bp in length
            *    71492    71591: gap of unknown length
            *    71592    73355: contig of 1764 bp in length
            *    73356    73455: gap of unknown length
            *    73456    75777: contig of 2322 bp in length
            *    75778    75877: gap of unknown length
            *    75878    78392: contig of 2515 bp in length
            *    78393    78492: gap of unknown length
            *    78493    81849: contig of 3357 bp in length
            *    81850    81949: gap of unknown length
            *    81950    84202: contig of 2253 bp in length
            *    84203    84302: gap of unknown length
            *    84303    86896: contig of 2594 bp in length
            *    86897    86996: gap of unknown length
            *    86997    90467: contig of 3471 bp in length
            *    90468    90567: gap of unknown length
            *    90568    93764: contig of 3197 bp in length
            *    93765    93864: gap of unknown length
            *    93865    96842: contig of 2978 bp in length
            *    96843    96942: gap of unknown length
            *    96943    99989: contig of 3047 bp in length
            *    99990   100089: gap of unknown length
            *   100090   103155: contig of 3066 bp in length
            *   103156   103255: gap of unknown length

  Query Match             14.2%;  Score 69.8;  DB 2;  Length 151358;
  Best Local Similarity   51.9%;  Pred. No. 0.0025;
  Matches  149;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy      84 CCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACC 143
```

```
                 || || || ||| | |  ||  | | |||  ||||  || | || || ||  |||   ||
Db    2498 CCCCCCCCCCCCCCCCNNCCCNCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCACCCCCCCCCC 2557

Qy     144 CCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCG 203
             |   | | || || || || || | ||||| | ||| | || ||| |   |    |||
Db    2558 TCTTTTTCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCTCTCCTCCCCCCCCCCCGGCG 2617

Qy     204 CGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCC 263
             |   ||| | |   |      | ||    |  || ||  | |||| | || ||  || |
Db    2618 GGCCCGGGGGGCGCGCTCNCCNTCCGCGCGGCCCCCCGGCCCCCGCGGCGGGGGGGGGGC 2677

Qy     264 GGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGG 323
           || | || || || |  | ||    || |       ||  |  || || || ||  |||
Db    2678 GGGGCGCGGCCGCGGGCGCGGGCGGGGGGCGCCGCGGGGGGCGGGGGGCGGGGCGGCGGC 2737

Qy     324 CGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGG 370
           |||  |  ||   |||   |  |   |  | |  || |  ||||
Db    2738 CGGGCGCGGCCCGCGGGCGCGCTCCGGCGGCCTGCCCCCCCGCGCGG 2784
```

RESULT 12
AC121227
LOCUS AC121227 174605 bp DNA linear HTG 24-JUL-2002
DEFINITION Rattus norvegicus clone CH230-103J5, *** SEQUENCING IN PROGRESS ***, 80 unordered pieces.
ACCESSION AC121227
VERSION AC121227.2 GI:21909120
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 174605)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,F., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174605)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 174605)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 19, 2002 this sequence version replaced gi:20806257.

```
---------- Genome Center
    Center: Baylor College of Medicine
    Center code: BCM
    Web site: http://www.hgsc.bcm.tmc.edu/
    Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
    Center project name: GYXK
    Center clone name: CH230-103J5
---------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 110654 bases at least Q40
    Consensus quality: 117178 bases at least Q30
    Consensus quality: 122597 bases at least Q20
----------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 80 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1007: contig of 1007 bp in length
*     1008     1107: gap of unknown length
*     1108     2450: contig of 1343 bp in length
*     2451     2550: gap of unknown length
*     2551     4232: contig of 1682 bp in length
*     4233     4332: gap of unknown length
*     4333     5794: contig of 1462 bp in length
*     5795     5894: gap of unknown length
*     5895     6962: contig of 1068 bp in length
*     6963     7062: gap of unknown length
*     7063     8247: contig of 1185 bp in length
*     8248     8347: gap of unknown length
*     8348     9503: contig of 1156 bp in length
*     9504     9603: gap of unknown length
*     9604    10846: contig of 1243 bp in length
*    10847    10946: gap of unknown length
*    10947    12170: contig of 1224 bp in length
*    12171    12270: gap of unknown length
*    12271    13571: contig of 1301 bp in length
*    13572    13671: gap of unknown length
*    13672    14731: contig of 1060 bp in length
*    14732    14831: gap of unknown length
*    14832    16136: contig of 1305 bp in length
*    16137    16236: gap of unknown length
*    16237    17769: contig of 1533 bp in length
*    17770    17869: gap of unknown length
*    17870    19119: contig of 1250 bp in length
*    19120    19219: gap of unknown length
*    19220    20422: contig of 1203 bp in length
*    20423    20522: gap of unknown length
*    20523    22121: contig of 1599 bp in length
```

```
*    22122    22221: gap of unknown length
*    22222    24265: contig of 2044 bp in length
*    24266    24365: gap of unknown length
*    24366    25970: contig of 1605 bp in length
*    25971    26070: gap of unknown length
*    26071    27332: contig of 1262 bp in length
*    27333    27432: gap of unknown length
*    27433    28627: contig of 1195 bp in length
*    28628    28727: gap of unknown length
*    28728    29885: contig of 1158 bp in length
*    29886    29985: gap of unknown length
*    29986    31804: contig of 1819 bp in length
*    31805    31904: gap of unknown length
*    31905    33339: contig of 1435 bp in length
*    33340    33439: gap of unknown length
*    33440    35142: contig of 1703 bp in length
*    35143    35242: gap of unknown length
*    35243    36736: contig of 1494 bp in length
*    36737    36836: gap of unknown length
*    36837    38640: contig of 1804 bp in length
*    38641    38740: gap of unknown length
*    38741    39914: contig of 1174 bp in length
*    39915    40014: gap of unknown length
*    40015    41570: contig of 1556 bp in length
*    41571    41670: gap of unknown length
*    41671    43016: contig of 1346 bp in length
*    43017    43116: gap of unknown length
*    43117    44156: contig of 1040 bp in length
*    44157    44256: gap of unknown length
*    44257    45989: contig of 1733 bp in length
*    45990    46089: gap of unknown length
*    46090    47638: contig of 1549 bp in length
*    47639    47738: gap of unknown length
*    47739    49218: contig of 1480 bp in length
*    49219    49318: gap of unknown length
*    49319    50429: contig of 1111 bp in length
*    50430    50529: gap of unknown length
*    50530    52925: contig of 2396 bp in length
*    52926    53025: gap of unknown length
*    53026    54637: contig of 1612 bp in length
*    54638    54737: gap of unknown length
*    54738    57055: contig of 2318 bp in length
*    57056    57155: gap of unknown length
*    57156    58371: contig of 1216 bp in length
*    58372    58471: gap of unknown length
*    58472    60149: contig of 1678 bp in length
*    60150    60249: gap of unknown length
*    60250    61548: contig of 1299 bp in length
*    61549    61648: gap of unknown length
*    61649    63896: contig of 2248 bp in length
*    63897    63996: gap of unknown length
*    63997    66103: contig of 2107 bp in length
*    66104    66203: gap of unknown length
*    66204    68167: contig of 1964 bp in length
*    68168    68267: gap of unknown length
*    68268    69807: contig of 1540 bp in length
*    69808    69907: gap of unknown length
*    69908    71931: contig of 2024 bp in length
*    71932    72031: gap of unknown length
*    72032    73191: contig of 1160 bp in length
*    73192    73291: gap of unknown length
*    73292    74767: contig of 1476 bp in length
*    74768    74867: gap of unknown length
*    74868    76389: contig of 1522 bp in length
*    76390    76489: gap of unknown length
*    76490    77915: contig of 1426 bp in length
*    77916    78015: gap of unknown length
*    78016    80236: contig of 2221 bp in length
*    80237    80336: gap of unknown length
*    80337    82748: contig of 2412 bp in length
*    82749    82848: gap of unknown length
*    82849    85565: contig of 2717 bp in length
*    85566    85665: gap of unknown length
*    85666    87488: contig of 1823 bp in length
*    87489    87588: gap of unknown length
*    87589    88625: contig of 1037 bp in length

  Query Match             13.9%;  Score 68.4;  DB 2;  Length 174605;
  Best Local Similarity   49.2%;  Pred. No. 0.0044;
  Matches  150;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy      3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
          | ||   |  | | ||||                |     |||  |  || || ||| |
Db  17159 CCCCCTCCCCCTTCCTCCNNNNNNNNNNNNNCCTNCCCNNNCCCCCCCCCCCCCCCCGCCC 17218

Qy     63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
          || ||||   || | || |  || || || ||| | |  || || | |||  |||| ||
Db  17219 CCCCCCCGCCGCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 17278

Qy    123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182
          | || || |   |||    |||| | | | |  || || |  ||  | || |  |||||
Db  17279 CCCCCNCCCCCCCCCCCCCGCCCCCCCCCCCCGCCCCCCCCCNCCCCGGGCGGCGGGCCCG 17338

Qy    183 TGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGT 242
          | || |          | ||||| |      |   |||  | |   ||   | |||
Db  17339 GGGCGGCGCGCCGGGGGGGCGCGGGCCCCGCCGCCCCGGCGCGCGGGGGGCGCGCCGGCG 17398

Qy    243 TCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGC 302
          ||||||   | |||     ||  | |  ||||  | | ||  |  |||  ||   || |
Db  17399 CCGCCGCCCCGCCGGGGCCGCGCGGGGGCGCGGGCGCGCGGCGGGCGGCCGCGCGCCGCC 17458

Qy    303 TGGTG 307
          || |
Db  17459 GGGCG 17463


RESULT 13
AC118303
LOCUS       AC118303               174031 bp    DNA     linear   HTG 18-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-173H17, *** SEQUENCING IN PROGRESS
            ***, 58 unordered pieces.
ACCESSION   AC118303
VERSION     AC118303.5  GI:21746412
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 174031)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
```

Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 174031)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (15-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 174031)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 14, 2002 this sequence version replaced gi:20330629.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GUJG
Center clone name: CH230-173H17
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 107398 bases at least Q40
Consensus quality: 119658 bases at least Q30
Consensus quality: 127269 bases at least Q20
----

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*        1     1024: contig of 1024 bp in length
*     1025     1124: gap of unknown length
*     1125     2167: contig of 1043 bp in length
*     2168     2267: gap of unknown length
*     2268     3414: contig of 1147 bp in length
*     3415     3514: gap of unknown length
*     3515     4714: contig of 1200 bp in length
*     4715     4814: gap of unknown length
*     4815     6353: contig of 1539 bp in length
*     6354     6453: gap of unknown length
*     6454     7915: contig of 1462 bp in length
*     7916     8015: gap of unknown length
*     8016     9359: contig of 1344 bp in length
*     9360     9459: gap of unknown length
*     9460    11082: contig of 1623 bp in length
*    11083    11182: gap of unknown length
*    11183    12369: contig of 1187 bp in length
*    12370    12469: gap of unknown length
*    12470    13930: contig of 1461 bp in length
*    13931    14030: gap of unknown length
*    14031    15253: contig of 1223 bp in length
*    15254    15353: gap of unknown length
*    15354    16696: contig of 1343 bp in length
*    16697    16796: gap of unknown length
*    16797    18403: contig of 1607 bp in length
*    18404    18503: gap of unknown length
*    18504    19776: contig of 1273 bp in length
*    19777    19876: gap of unknown length
*    19877    21648: contig of 1772 bp in length
*    21649    21748: gap of unknown length
*    21749    23094: contig of 1346 bp in length
*    23095    23194: gap of unknown length
*    23195    25092: contig of 1898 bp in length
*    25093    25192: gap of unknown length
*    25193    26802: contig of 1610 bp in length
*    26803    26902: gap of unknown length
*    26903    28726: contig of 1824 bp in length
*    28727    28826: gap of unknown length
*    28827    31064: contig of 2238 bp in length
*    31065    31164: gap of unknown length
*    31165    33254: contig of 2090 bp in length
*    33255    33354: gap of unknown length
*    33355    35763: contig of 2409 bp in length
*    35764    35863: gap of unknown length
*    35864    36989: contig of 1126 bp in length
*    36990    37089: gap of unknown length
*    37090    38703: contig of 1614 bp in length
*    38704    38803: gap of unknown length
*    38804    40934: contig of 2131 bp in length
*    40935    41034: gap of unknown length
*    41035    42978: contig of 1944 bp in length
*    42979    43078: gap of unknown length
*    43079    45214: contig of 2136 bp in length
*    45215    45314: gap of unknown length
*    45315    47780: contig of 2466 bp in length
*    47781    47880: gap of unknown length
*    47881    50209: contig of 2329 bp in length
*    50210    50309: gap of unknown length
*    50310    52364: contig of 2055 bp in length
*    52365    52464: gap of unknown length
*    52465    54739: contig of 2275 bp in length
*    54740    54839: gap of unknown length
*    54840    56920: contig of 2081 bp in length
*    56921    57020: gap of unknown length
*    57021    59676: contig of 2656 bp in length
*    59677    59776: gap of unknown length
*    59777    62058: contig of 2282 bp in length
*    62059    62158: gap of unknown length
*    62159    64481: contig of 2323 bp in length
*    64482    64581: gap of unknown length
*    64582    67737: contig of 3156 bp in length
*    67738    67837: gap of unknown length
*    67838    69856: contig of 2019 bp in length
*    69857    69956: gap of unknown length
*    69957    72698: contig of 2742 bp in length
*    72699    72798: gap of unknown length
*    72799    75900: contig of 3102 bp in length
*    75901    76000: gap of unknown length
*    76001    78614: contig of 2614 bp in length
*    78615    78714: gap of unknown length
*    78715    81699: contig of 2985 bp in length
*    81700    81799: gap of unknown length
*    81800    85909: contig of 4110 bp in length
*    85910    86009: gap of unknown length
*    86010    89796: contig of 3787 bp in length
*    89797    89896: gap of unknown length
*    89897    93635: contig of 3739 bp in length
*    93636    93735: gap of unknown length
*    93736    97658: contig of 3923 bp in length
*    97659    97758: gap of unknown length
*    97759   101045: contig of 3287 bp in length
```

```
            *   101046   101145: gap of unknown length
            *   101146   105352: contig of 4207 bp in length
            *   105353   105452: gap of unknown length
            *   105453   109545: contig of 4093 bp in length
            *   109546   109645: gap of unknown length
            *   109646   114476: contig of 4831 bp in length
            *   114477   114576: gap of unknown length
            *   114577   118745: contig of 4169 bp in length
            *   118746   118845: gap of unknown length
            *   118846   122388: contig of 3543 bp in length
            *   122389   122488: gap of unknown length
            *   122489   128899: contig of 6411 bp in length
            *   128900   128999: gap of unknown length
            *   129000   134742: contig of 5743 bp in length
            *   134743   134842: gap of unknown length
            *   134843   139853: contig of 5011 bp in length

 Query Match             13.8%;  Score 68;  DB 2;  Length 174031;
 Best Local Similarity   49.5%;  Pred. No. 0.0052;
 Matches  158;  Conservative    0;  Mismatches  159;  Indels    2;  Gaps    1;

Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTC 76
           | ||  |||  |       ||      ||      || || ||    || | ||    | |
Db   24645 CCCCCNCCCCCCNCCCCCCCCNNCNNCCNNCNCCCCCCCCCCCNCGCCCCCNCCCCCCCNC 24704

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
             ||    || || || ||| | |  || || | |||  ||||  || | ||         |
Db   24705 CCCCNCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNNNNNCNC 24764

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | |  | || || ||  | |     | || | | | || | | |
Db   24765 CCNCCCCCCCCCGCCCCNCCCCCCCCCCCCGCCCCNNNNCCCGGCGGCGGC  GGGGCNG 24824

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
             | ||| | | ||| ||  |  |  | |   ||   ||||  ||   |  |  ||   |
Db   24825 CGGGGCGGGGGNGGGGGCGGGGGGCGGCGGCGGGGGGGGGCGCCGGGGGGGGGGGGCCGG 24884

Qy     257 GCCG--GCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGG 314
             |||  ||||| |  | |||   | ||||| |  ||||  ||    || ||  |   |  |
Db   24885 CCCGGCGCCGGGGGCCGGCGCGCGGGGGGGCGGGGGCGGGGGCGCGCGGGCGCNCGGCG 24944

Qy     315 CGCAACGGGCGGCGGACGC 333
           |||    |||||||  | |
Db   24945 CGCGGGCGGCGGCGCGCCC 24963


RESULT 14
AC106398
LOCUS       AC106398               68787 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-137B1, *** SEQUENCING IN PROGRESS
            ***, 37 unordered pieces.
ACCESSION   AC106398
VERSION     AC106398.2  GI:21735179
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 68787)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68787)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 68787)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:18138919.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GKUW
                Center clone name: CH230-137B1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 35643 bases at least Q40
                Consensus quality: 39014 bases at least Q30
                Consensus quality: 41785 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1070: contig of 1070 bp in length
            *     1071     1170: gap of unknown length
            *     1171     2678: contig of 1508 bp in length
            *     2679     2778: gap of unknown length
            *     2779     4068: contig of 1290 bp in length
```

```
*     4069     4168: gap of unknown length
*     4169     5608: contig of 1440 bp in length
*     5609     5708: gap of unknown length
*     5709     7127: contig of 1419 bp in length
*     7128     7227: gap of unknown length
*     7228     8577: contig of 1350 bp in length
*     8578     8677: gap of unknown length
*     8678     9793: contig of 1116 bp in length
*     9794     9893: gap of unknown length
*     9894    10907: contig of 1014 bp in length
*    10908    11007: gap of unknown length
*    11008    12349: contig of 1342 bp in length
*    12350    12449: gap of unknown length
*    12450    13548: contig of 1099 bp in length
*    13549    13648: gap of unknown length
*    13649    14730: contig of 1082 bp in length
*    14731    14830: gap of unknown length
*    14831    16251: contig of 1421 bp in length
*    16252    16351: gap of unknown length
*    16352    18478: contig of 2127 bp in length
*    18479    18578: gap of unknown length
*    18579    20188: contig of 1610 bp in length
*    20189    20288: gap of unknown length
*    20289    21551: contig of 1263 bp in length
*    21552    21651: gap of unknown length
*    21652    23305: contig of 1654 bp in length
*    23306    23405: gap of unknown length
*    23406    25560: contig of 2155 bp in length
*    25561    25660: gap of unknown length
*    25661    27283: contig of 1623 bp in length
*    27284    27383: gap of unknown length
*    27384    29246: contig of 1863 bp in length
*    29247    29346: gap of unknown length
*    29347    31576: contig of 2230 bp in length
*    31577    31676: gap of unknown length
*    31677    33313: contig of 1637 bp in length
*    33314    33413: gap of unknown length
*    33414    35876: contig of 2463 bp in length
*    35877    35976: gap of unknown length
*    35977    37553: contig of 1577 bp in length
*    37554    37653: gap of unknown length
*    37654    39176: contig of 1523 bp in length
*    39177    39276: gap of unknown length
*    39277    40741: contig of 1465 bp in length
*    40742    40841: gap of unknown length
*    40842    42390: contig of 1549 bp in length
*    42391    42490: gap of unknown length
*    42491    44654: contig of 2164 bp in length
*    44655    44754: gap of unknown length
*    44755    46171: contig of 1417 bp in length
*    46172    46271: gap of unknown length
*    46272    48302: contig of 2031 bp in length
*    48303    48402: gap of unknown length
*    48403    50074: contig of 1672 bp in length
*    50075    50174: gap of unknown length
*    50175    52341: contig of 2167 bp in length
*    52342    52441: gap of unknown length
*    52442    54601: contig of 2160 bp in length
*    54602    54701: gap of unknown length
*    54702    57006: contig of 2305 bp in length
*    57007    57106: gap of unknown length
*    57107    59715: contig of 2609 bp in length
*    59716    59815: gap of unknown length
*    59816    62344: contig of 2529 bp in length
*    62345    62444: gap of unknown length
*    62445    66001: contig of 3557 bp in length
*    66002    66101: gap of unknown length
*    66102    68787: contig of 2686 bp in length.
FEATURES             Location/Qualifiers
     source          1. .68787
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-137B1"
```

```
BASE COUNT    18229 a  14391 c  13550 g  17475 t   5142 others
ORIGIN

  Query Match            13.7%;  Score 67.4;  DB 2;  Length 68787;
  Best Local Similarity  54.1%;  Pred. No. 0.0071;
  Matches  157;  Conservative    0;  Mismatches  132;  Indels    1;  Gaps    1;

Qy      16 TCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCT 75
           ||  | ||||  |  |       | |||||  |  || || ||  | || ||||  | ||
Db   31191 TCCTCNCCCCCCCCCTGCTCGTCTCCTCCCCCCCCCCCCCCCCTCCCCCCCCCCCCTCCT 31250

Qy      76 CTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAA 135
           | || || || || || ||| | |   | || | |||||||;||  || | || || |
Db   31251 CCCCCCTCCCCCCCCCCCCCCCCCCCCCGCTCCCCCCCCCTTCCCCCCCCCCCCCCCTCTCC 31310

Qy     136 CCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGT 195
           |||   ||||| | | | |    ||| || || |  |||     | | | ||  | ||
Db   31311 CCCCCCCCCCCTCCCCCCCTCGTTCCCCCCCCCCCCGCGCGGTCTCGGCGGCGCGCCCGCG 31370

Qy     196 CTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGC 255
             ||||   | ||||| ||||  |     ||     ||| ||   |  ||||  ||
Db   31371 GCGCGTCGCGGCGGCGCGCCCGCCGGCGGCCGGGGCCGCCGCCGCG-GGGGGGCTGCCGC 31429

Qy     256 GGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGCTGG 305
           ||||||  |    ||    || ||||||     ||||  ||  |||   ||
Db   31430 CGCCGGCGGCGCCGCCTGCGCCGTGGGGCGTCGGGCGGCGCTGCGCGGGG 31479


RESULT 15
AC126730
LOCUS       AC126730               130899 bp    DNA     linear   HTG 23-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-6H4, *** SEQUENCING IN PROGRESS ***,
            69 unordered pieces.
ACCESSION   AC126730
VERSION     AC126730.2  GI:21886838
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 130899)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
```

Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 130899)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 130899)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 17, 2002 this sequence version replaced gi:21716710.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GBWP
Center clone name: CH230-6H4
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 71039 bases at least Q40
Consensus quality: 77045 bases at least Q30
Consensus quality: 80907 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1027: contig of 1027 bp in length
* 1028 1127: gap of unknown length
* 1128 2272: contig of 1145 bp in length
* 2273 2372: gap of unknown length
* 2373 3398: contig of 1026 bp in length
* 3399 3498: gap of unknown length
* 3499 4802: contig of 1304 bp in length
* 4803 4902: gap of unknown length
* 4903 5993: contig of 1091 bp in length
* 5994 6093: gap of unknown length
* 6094 7244: contig of 1151 bp in length
* 7245 7344: gap of unknown length
* 7345 8880: contig of 1536 bp in length
* 8881 8980: gap of unknown length
* 8981 9987: contig of 1007 bp in length
* 9988 10087: gap of unknown length
* 10088 11342: contig of 1255 bp in length
* 11343 11442: gap of unknown length
* 11443 12560: contig of 1118 bp in length
* 12561 12660: gap of unknown length
* 12661 13884: contig of 1224 bp in length
* 13885 13984: gap of unknown length
* 13985 15064: contig of 1080 bp in length
* 15065 15164: gap of unknown length
* 15165 16546: contig of 1382 bp in length
* 16547 16646: gap of unknown length
* 16647 18513: contig of 1867 bp in length
* 18514 18613: gap of unknown length
* 18614 19705: contig of 1092 bp in length
* 19706 19805: gap of unknown length
* 19806 21270: contig of 1465 bp in length
* 21271 21370: gap of unknown length
* 21371 22600: contig of 1230 bp in length
* 22601 22700: gap of unknown length
* 22701 23741: contig of 1041 bp in length
* 23742 23841: gap of unknown length
* 23842 24945: contig of 1104 bp in length
* 24946 25045: gap of unknown length
* 25046 27258: contig of 2213 bp in length
* 27259 27358: gap of unknown length
* 27359 28617: contig of 1259 bp in length
* 28618 28717: gap of unknown length
* 28718 29892: contig of 1175 bp in length
* 29893 29992: gap of unknown length
* 29993 31396: contig of 1404 bp in length
* 31397 31496: gap of unknown length
* 31497 32504: contig of 1008 bp in length
* 32505 32604: gap of unknown length
* 32605 34000: contig of 1396 bp in length
* 34001 34100: gap of unknown length
* 34101 35857: contig of 1757 bp in length
* 35858 35957: gap of unknown length
* 35958 37377: contig of 1420 bp in length
* 37378 37477: gap of unknown length
* 37478 38497: contig of 1020 bp in length
* 38498 38597: gap of unknown length
* 38598 39693: contig of 1096 bp in length
* 39694 39793: gap of unknown length
* 39794 42502: contig of 2709 bp in length
* 42503 42602: gap of unknown length
* 42603 44423: contig of 1821 bp in length
* 44424 44523: gap of unknown length
* 44524 46230: contig of 1707 bp in length
* 46231 46330: gap of unknown length
* 46331 47759: contig of 1429 bp in length
* 47760 47859: gap of unknown length
* 47860 49208: contig of 1349 bp in length
* 49209 49308: gap of unknown length
* 49309 50360: contig of 1052 bp in length
* 50361 50460: gap of unknown length
* 50461 52257: contig of 1797 bp in length
* 52258 52357: gap of unknown length
* 52358 54147: contig of 1790 bp in length
* 54148 54247: gap of unknown length
* 54248 55402: contig of 1155 bp in length
* 55403 55502: gap of unknown length
* 55503 57008: contig of 1506 bp in length
* 57009 57108: gap of unknown length
* 57109 58547: contig of 1439 bp in length
* 58548 58647: gap of unknown length
* 58648 60159: contig of 1512 bp in length
* 60160 60259: gap of unknown length
* 60260 61728: contig of 1469 bp in length
* 61729 61828: gap of unknown length
* 61829 64395: contig of 2567 bp in length
* 64396 64495: gap of unknown length
* 64496 66096: contig of 1601 bp in length
* 66097 66196: gap of unknown length
* 66197 67712: contig of 1516 bp in length
* 67713 67812: gap of unknown length
* 67813 68997: contig of 1185 bp in length
* 68998 69097: gap of unknown length
* 69098 70867: contig of 1770 bp in length
* 70868 70967: gap of unknown length
* 70968 72129: contig of 1162 bp in length
* 72130 72229: gap of unknown length

```
          *    72230    75006: contig of 2777 bp in length
          *    75007    75106: gap of unknown length
          *    75107    76249: contig of 1143 bp in length
          *    76250    76349: gap of unknown length
          *    76350    79053: contig of 2704 bp in length
          *    79054    79153: gap of unknown length
          *    79154    81129: contig of 1976 bp in length
          *    81130    81229: gap of unknown length
          *    81230    83849: contig of 2620 bp in length
          *    83850    83949: gap of unknown length
          *    83950    86640: contig of 2691 bp in length

  Query Match             13.7%;  Score 67.4;  DB 2;  Length 130899;
  Best Local Similarity   46.3%;  Pred. No. 0.0068;
  Matches  158;  Conservative    0;  Mismatches  183;  Indels    0;  Gaps    0;

Qy      36 CCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCC 95
           ||    | |||  |  || || ||    || ||||        || |  ||  | || |||
Db   89941 CCCCCCCCCCCCCNCCCNCCCCCNNNCCCCCCCCCCCCCCNNNNCCCCCCCCCNCCCCCCCC 90000

Qy      96 ACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCC 155
             |  || | |||  ||||  |    || || |   |||   ||||| | | | ||
Db   90001 NAAGCCCNCCCCCCCCCCCCCCCCCCCNNNCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCC 90060

Qy     156 TCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCC 215
            || ||| |  |       | || |       |   || | |  |   | |   ||| |
Db   90061 CCCCCCGGCGACNNNNGCGCCCGNCGCGCCGGCGNGCCGGCGCGCCGGGGGCGGGGGGGG 90120

Qy     216 AGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGG 275
            ||    | |   ||   ||| | ||   ||||| |||   | |||||||| | |  |
Db   90121 GGCGCCGGGGGCGGGGGGGGGGCGCGCGGGCCGCGGGGNGCGGCGGCCGGGGCGGCGGCC 90180

Qy     276 CAGTGGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTG 335
           | |  ||   |  |||| | | ||||   | |  |  ||    ||| ||  |   |
Db   90181 CCGGCGGNCCGCGGGCGNCCGCGCGGGNCGNGCGGCGGCGCGCGCGGNCGCGGNCGGGGC 90240

Qy     336 CTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGGTTGAG 376
           |    | || |  |  |  |  |     |  |||||  | |
Db   90241 CCCCCGCGGCGGGCGGGNGGCGCGCGGNCGGGCGGGGGGGG 90281


Search completed: November 17, 2002, 16:19:03
Job time : 3759 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: November 18, 2002, 21:53:35 ; Search time 269 Seconds
(without alignments)
4118.897 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agctccgagcgtcattctcc.....agccaattttggngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	489	99.4	492	21	AAZ29165	Corn extragenic su
	2	165.8	33.7	1418	21	AAZ29168	Partial cDNA inser
	3	69	14.0	1103	21	AAZ29167	Assembled cDNA con
	4	64.4	13.1	117213	19	AAV62176	HSV-2 strain SB5 C
c	5	59	12.0	600	24	ABQ52496	Oligonucleotide fo
	6	59	12.0	600	24	ABQ52497	Oligonucleotide fo
	7	58.6	11.9	542	23	AAS67812	DNA encoding novel
c	8	58.2	11.8	1266	20	AAX09011	Brn-3a polynucleot
c	9	58.2	11.8	1266	21	AAA29007	Murine transcripti
	10	56.4	11.5	9953	23	ABK42449	Genomic sequence #
	11	56.4	11.5	11189	23	ABK42448	Genomic sequence #
c	12	56	11.4	434	22	AAI84343	Human polynucleoti
c	13	56	11.4	1286	24	ABI99656	Mouse ischaemic co
	14	55.8	11.3	853	23	AAS67813	DNA encoding novel
c	15	55	11.2	1045	24	ABI99807	Mouse ischaemic co
c	16	55	11.2	1327	24	ABQ68452	Listeria monocytog
	17	54.4	11.1	246	24	ABL51825	Ser-Pro-Pro-Pro ge
c	18	54.4	11.1	320	21	AAA38183	Primer used in the
c	19	54.4	11.1	349	22	AAI84504	Human polynucleoti
c	20	54.4	11.1	851	24	ABQ17074	Oligonucleotide fo
	21	54.4	11.1	851	24	ABQ17075	Oligonucleotide fo
	22	54.4	11.1	2188	20	AAZ77506	Human ovarian tumo
c	23	53.8	10.9	318	21	AAA38184	Primer used in the
c	24	53.8	10.9	344	22	AAI82007	Human polynucleoti
c	25	53.8	10.9	2721	12	AAQ10212	BamHI J-I fragment
c	26	53.8	10.9	2721	12	AAQ10543	BamHI J-I fragment
c	27	53.8	10.9	8438	15	AAQ73500	DNA encoding Pseud
c	28	53.6	10.9	320	21	AAA38185	Primer used in the
	29	52.8	10.7	12001	16	AAQ76213	HSV L/ST region.
c	30	52.6	10.7	320	21	AAA38186	Primer used in the
c	31	52.6	10.7	1336	20	AAZ16779	Human gene express
c	32	52.6	10.7	1336	21	AAA01591	Human colon cancer
c	33	52.2	10.6	1478	24	ABQ30124	Oligonucleotide fo
	34	52.2	10.6	1478	24	ABQ30125	Oligonucleotide fo
	35	51.8	10.5	1614	22	AAH26499	Human low density
	36	51.8	10.5	12425	22	AAH26495	Human low density
c	37	51.4	10.4	237	21	AAA81807	N. meningitidis pa
	38	51.2	10.4	1337	20	AAZ17263	Human gene express
	39	51.2	10.4	5025	24	ABK70305	Human lung cancer
c	40	51	10.4	463	22	AAC84589	Rice magnesium che
	41	51	10.4	1272	20	AAX09010	Brn-3a polynucleot
	42	51	10.4	1272	21	AAA29006	Human transcriptio
c	43	50.6	10.3	168	24	ABN69641	Streptococcus poly
	44	50.6	10.3	1008	22	AAH21004	Bovine-derived DNA
c	45	50.6	10.3	61710	22	AAK83782	Human immune/haema

ALIGNMENTS

RESULT 1
AAZ29165
ID AAZ29165 standard; DNA; 492 BP.
XX
AC AAZ29165;
XX
DT 21-FEB-2000 (first entry)
XX
DE Corn extragenic suppressor protein cDNA.
XX
KW Clone cdt2c.pk003.b20; corn extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW genetic mapping; animal feed; ds.
XX
OS Zea mays.
XX

FH	Key	Location/Qualifiers
FT	CDS	71..492
FT		/*tag= a
FT		/transl_except= (pos:101..103, aa: Asp)
FT		/transl_except= (pos:422..492, aa: Xaa)
FT		/note= "Xaa= amino acids 118-338 of corn extragenic
FT		suppressor protein"
FT	unsure	350..352
FT		/*tag= b
FT		/note= "Encodes Arg"

XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX

PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR P-PSDB; AAY44230.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 7; Page 37; 60pp; English.
XX
CC The present sequence is the cDNA insert in clone cdt2c.pk003.b20 encoding
CC corn extragenic suppressor protein. The cDNA clone was isolated from corn
CC developing tassle. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising this cDNA. This is done to increase levels of
CC free phosphate and decrease phytic acid levels in developing seeds or in
CC the construction of transgenic plants with altered levels of the enzyme.
CC The phytic acid biosynthetic enzymes are used to raise antibodies, for
CC enzyme detection in cells and cell extracts. The polynucleotides also
CC serve as a source of probes and primers, for genetic mapping, as markers,
CC and to isolate homologous sequences from other species. This is
CC commercially used for animal feed.
XX
SQ Sequence 492 BP; 88 A; 169 C; 144 G; 88 T; 3 other;

Query Match 99.4%; Score 489; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 9.7e-91;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 AGCTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGCTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGA 60

Qy      61 CTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGG 120

Qy     121 CCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCC 180

Qy     181 CGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTC 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTC 240

Qy     241 GTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCG 300

Qy     301 GCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTT 360

Qy     361 CCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAG 420

Qy     421 AAGAAGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AAGAAGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCCAATT 480

Qy     481 TTGGNGAAGAGA 492
          ||||||||||||
Db     481 TTGGNGAAGAGA 492
```

RESULT 2

AAZ29168
ID AAZ29168 standard; DNA; 1418 BP.
XX
AC AAZ29168;
XX
DT 21-FEB-2000 (first entry)
XX
DE Partial cDNA insert encoding wheat extragenic suppressor protein.
XX
KW Clone wlm0.pk0010.f6; wheat extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW genetic mapping; animal feed; ds.
XX
OS Triticum aestivum.
XX
FH Key Location/Qualifiers
FT CDS 2..976
FT /*tag= a
FT /product= "Wheat extragenic suppressor protein"
FT /note= "Partial sequence"
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR P-PSDB; AAY44233.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 7; Page 40-41; 60pp; English.
XX
CC The present sequence is the cDNA insert in clone wlm0.pk0010.f6 encoding
CC a portion of wheat extragenic suppressor protein. The cDNA clone was
CC isolated from wheat seedlings. The level of expression of this phytic
CC acid biosynthetic enzyme can be altered by transforming a host cell with
CC the chimeric gene comprising this cDNA. This is done to increase levels
CC of free phosphate and decrease phytic acid levels in developing seeds or
CC in the construction of transgenic plants with altered levels of the
CC enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. The
CC polynucleotides also serve as a source of probes and primers, for genetic
CC mapping, as markers, and to isolate homologous sequences from other
CC species. This is commercially used for animal feed.
XX
SQ Sequence 1418 BP; 393 A; 307 C; 386 G; 332 T; 0 other;

Query Match 33.7%; Score 165.8; DB 21; Length 1418;
Best Local Similarity 69.7%; Pred. No. 6.9e-25;
Matches 251; Conservative 0; Mismatches 105; Indels 4; Gaps 2;

```
Qy     136 CCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGT 195
           |||| ||| || |      | |  || |||     ||  |||  ||    || | |
Db      22 CCCTTACCACCTCCGCTCGCCCTCCCTCCTGGCAACCTTCTCCTCCTCCGCGGCGGGTCG 81

Qy     196 CTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGC 255
             ||| ||  | | |||    |||| |  ||||| | ||||||  | | |  | ||
Db      82 GGCCTGCGGGATAGCGGGCCGTTGGATGGGCTCGGTTCGAGCCTCGCCCTCTGAGGCGGG 141

Qy     256 GGCCGGCCGGAGAGCTGCGGCAGT---GGGGGAGTTGGCGACGGAGCGGCTGGTGGAGGT 312
           || | |   |  |||||||| |    || || | |||  | ||||||||||||| |||
Db     142 GGGCTGGGCGGTGGCTGCGGCGGGTAAGGAGGGGGTGGACATGGAGCGGCTGGTGGCGGT 201
```

```
Qy     313 GGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANCAAGTACTTCCGCCAGCGGGT 372
           ||||||  | ||||||| || || ||||||||||||| |||||||||| | |||||  |
Db     202 GGCGCAGAGCGCGGCGGATGCGGCGGGGGAGGTGCTCAGGAAGTACTTCAGGCAGCGCTT 261

Qy     373 TGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAGATAGAAGAAGCAGAAG 432
           |||||||||||||||||||||||||||||| || || || || ||||| ||||||||||
Db     262 CGAGATCATCGACAAAGAGGACCACAGTCCCGTCACGATCGCTGATAG-AGAAGCAGAAG 320

Qy     433 AAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGC AATTTTGGNGAAGAGA 492
           ||||||||   ||||| |||||||||||||| |||||||| ||   ||| || || ||||
Db     321 AAGCAATGACCTCAGTCATACTGAAGAGCTTTCCTACTCATGCTGTTTTCGGTGAGGAGA 380
```

RESULT 3
AAZ29167
ID AAZ29167 standard; DNA; 1103 BP.
XX
AC AAZ29167;
XX
DT 21-FEB-2000 (first entry)
XX
DE Assembled cDNA contig encoding soybean extragenic suppressor protein.
XX
KW Clone s12.pk122.p24; clone src3c.pk013.g15; clone sf11n.pk003.g19;
KW soybean extragenic suppressor protein; phytic acid biosynthetic enzyme;
KW transformation; chimeric gene; antibody; free phosphate;
KW phytic acid level; transgenic plant; enzyme detection; genetic mapping;
KW animal feed; ds.
XX
OS Glycine max.
XX
FH Key   Location/Qualifiers
FT CDS   9..896
FT   /*tag= a
FT   /product= "Soybean extragenic suppressor protein"
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR P-PSDB; AAY44232.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 7; Page 39; 60pp; English.
XX
CC The present sequence is a contig assembled from portions of the cDNA
CC inserts in clones s12.pk122.p24, src3c.pk013.g15, and sf11n.pk003.g19,
CC encoding a soybean extragenic suppressor protein. The level of expression
CC of this phytic acid biosynthetic enzyme can be altered by transforming a
CC host cell with the chimeric gene comprising this cDNA. This is done to
CC increase levels of free phosphate and decrease phytic acid levels in
CC developing seeds or in the construction of transgenic plants with altered
CC levels of the enzyme. The phytic acid biosynthetic enzymes are used to
CC raise antibodies, for enzyme detection in cells and cell extracts. The
CC polynucleotides also serve as a source of probes and primers, for genetic
CC mapping, as markers, and to isolate homologous sequences from other
CC species. This is commercially used for animal feed.
XX
SQ Sequence 1103 BP; 307 A; 238 C; 236 G; 322 T; 0 other;

Query Match 14.0%; Score 69; DB 21; Length 1103;
Best Local Similarity 62.2%; Pred. No. 3.4e-05;

Matches 122; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

```
Qy     297 AGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGAGGTGCTCANGAAGT 356
           | ||  | |  ||||| |   |    |  || || ||||| || || ||  ||   || |
Db     106 ATCGCTTCGCCGAGGTCGGTAACAAAGTCGCCGATGCTGCCGGAGAAGTTATCCGCAAAT 165

Qy     357 ACTTCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACCACAGTCCTGTTACAATTGCAG 416
           ||||| |  |     | ||  | ||  ||||| | || | |||||| || || |||||||
Db     166 ACTTCAGAAAAAACTTCGACGTTATTCACAAACATGATCTCAGTCCAGTAACCATTGCAG 225

Qy     417 ATAGAAGAAGCAGAAGAAGCAATGGNGTCAGTTATACTGAAGAGCTTCCCTACTCAAGCC 476
           ||  ||   || || || || ||||  ||| | |||||  | |  |||||| |||| |||
Db     226 AT-CAATCTGCTGAGGAGGCTATGGTTTCAATCATACTAGACAATTTCCCTTCTCATGCC 284

Qy     477 AATTTTGGNGAAGAGA 492
           | ||  || || || |
Db     285 ATTTACGGAGAGGAAA 300
```

RESULT 4
AAV62176
ID AAV62176 standard; DNA; 117213 BP.
XX
AC AAV62176;
XX
DT 13-JAN-1999 (first entry)
XX
DE HSV-2 strain SB5 Contig ID 15 DNA sequence.
XX
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor; ss.
XX
OS Herpes simplex virus type 2.
XX
FH Key   Location/Qualifiers
FT CDS   755..1297
FT   /*tag= a
FT   /product= "ORF#1 protein"
FT   /note= "encoded protein shown in AAW72170"
FT CDS   1170..2174
FT   /*tag= b
FT   /product= "ORF#2 protein"
FT   /note= "encoded protein shown in AAW72171"
FT CDS   2229..2930
FT   /*tag= c
FT   /product= "ORF#3 protein"
FT   /note= "encoded protein shown in AAW72172"
FT CDS   complement (3130..3735)
FT   /*tag= d
FT   /product= "ORF#4 protein"
FT   /note= "encoded protein shown in AAW72173"
FT CDS   complement (3802..6447)
FT   /*tag= e
FT   /product= "ORF#5 protein"
FT   /note= "encoded protein shown in AAW72174"
FT CDS   6017..8482
FT   /*tag= f
FT   /product= "ORF#6g protein"
FT   /note= "encoded protein shown in AAW72250"
FT CDS   6026..8482
FT   /*tag= g
FT   /product= "ORF#6f protein"
FT   /note= "encoded protein shown in AAW72249"
FT CDS   6065..8482
FT   /*tag= h
FT   /product= "ORF#6e protein"
FT   /note= "encoded protein shown in AAW72248"
FT CDS   6167..8482
FT   /*tag= i
FT   /product= "ORF#6d protein"
FT   /note= "encoded protein shown in AAW72247"
FT CDS   6296..8482
FT   /*tag= j

```
FT                   /product= "ORF#6c protein"
FT                   /note= "encoded protein shown in AAW72246"
FT   CDS             6326..8482
FT                   /*tag= k
FT                   /product= "ORF#6b protein"
FT                   /note= "encoded protein shown in AAW72245"
FT   CDS             6446..8482
FT                   /*tag= l
FT                   /product= "ORF#6a protein"
FT                   /transl_except= (pos: 7400..7402, aa: Ala-Ala)
FT                   /transl_except= (pos: 7481..7486, aa: Ile)
FT                   /note= "encoded protein shown in AAW72244"
FT   CDS             8457..9347
FT                   /*tag= m
FT                   /product= "ORF#7 protein"
FT                   /note= "encoded protein shown in AAW72175"
FT   CDS             complement (9604..11855)
FT                   /*tag= n
FT                   /transl_except= (pos: 11635..11636, aa: Ala)
FT                   /product= "ORF#8 protein"
FT                   /note= "encoded protein showr in AAW72176"
FT   CDS             complement (11905..14508)
FT                   /*tag= o
FT                   /product= "ORF#9b protein"
FT                   /note= "encoded protein shown in AAW72222"
FT   CDS             complement (11905..14520)
FT                   /*tag= p
FT                   /product= "ORF#9a protein"
FT                   /note= "encoded protein shown in AAW72223"
FT   CDS             14399..15802
FT                   /*tag= q
FT                   /product= "ORF#10 protein"
FT                   /note= "encoded protein shown in AAW72177"
FT   CDS             complement (15996..16286)
FT                   /*tag= r
FT                   /product= "ORF#11 protein"
FT                   /note= "encoded protein shown in AAW72178"
FT   CDS             complement (16202..18064)
FT                   /*tag= s
FT                   /product= "ORF#12 protein"
FT                   /note= "encoded protein shown in AAW72179"
FT   CDS             complement (18105..19661)
FT                   /*tag= t
FT                   /product= "ORF#13 protein"
FT                   /note= "encoded protein shown in AAW72180"
FT   CDS             complement (19415..20074)
FT                   /*tag= u
FT                   /product= "ORF#14 protein"
FT                   /note= "encoded protein shown in AAW72181"
FT   CDS             20155..21453
FT                   /*tag= v
FT                   /product= "ORF#15 protein"
FT                   /note= "encoded protein shown in AAW72182"
FT   CDS             complement (21326..22291)
FT                   /*tag= w
FT                   /product= "ORF#16 protein"
FT                   /note= "encoded protein shown in AAW72183"
FT   CDS             complement (22546..24654)
FT                   /*tag= x
FT                   /product= "ORF#17 protein"
FT                   /note= "encoded protein shown in AAW72184"
FT   CDS             24684..25955
FT                   /*tag= y
FT                   /product= "ORF#18 protein"
FT                   /note= "encoded protein shown in AAW72185"
FT   CDS             complement (26295..27251)
FT                   /*tag= z
FT                   /product= "ORF#19 protein"
FT                   /note= "encoded protein shown in AAW72186"
FT   CDS             complement (27630..31754)
FT                   /*tag= aa
FT                   /product= "ORF#20b protein"
FT                   /note= "encoded protein shown in AAW72225"
FT   CDS             complement (27630..31784)
FT                   /*tag= ab
FT                   /product= "ORF#20a protein"
FT                   /note= "encoded protein shown in AAW72224"
FT   CDS             complement (32067..32735)
FT                   /*tag= ac
FT                   /product= "ORF#21 protein"
FT                   /note= "encoded protein shown in AAW72187"
FT   CDS             33140..34984
FT                   /*tag= ad
FT                   /product= "ORF#22a protein"
FT                   /note= "encoded protein shown in AAW72226"
FT   CDS             33386..34984
FT                   /*tag= ae
FT                   /product= "ORF#22b protein"
FT                   /note= "encoded protein shown in AAW72227"
FT   CDS             complement (35205..37721)
FT                   /*tag= af
FT                   /product= "ORF#23 protein"
FT                   /note= "encoded protein shown in AAW72188"
FT   CDS             complement (38058..39188)
FT                   /*tag= ag
FT                   /product= "ORF#24 protein"
FT                   /note= "encoded protein shown in AAW72189"
FT   CDS             39090..39935
FT                   /*tag= ah
FT                   /product= "ORF#25 protein"
FT                   /note= "encoded protein shown in AAW72190"
FT   CDS             40216..41973
FT                   /*tag= ai
FT                   /product= "ORF#26 protein"
FT                   /note= "encoded protein shown in AAW72191"
FT   CDS             42206..44178
FT                   /*tag= aj
FT                   /transl_except= (pos: 44063..44064, aa: Lys)
FT                   /product= "ORF#27 protein"
FT                   /note= "encoded protein shown in AAW72192"
FT   CDS             complement (44853..47297)
FT                   /*tag= ak
FT                   /product= "ORF#28 protein"
FT                   /note= "encoded protein shown in AAW72193"
FT   CDS             47122..47338
FT                   /*tag= al
FT                   /product= "ORF#29 protein"
FT                   /note= "encoded protein shown in AAW72194"
FT   CDS             complement (47305..49662)
FT                   /*tag= am
FT                   /product= "ORF#30 protein"
FT                   /note= "encoded protein shown in AAW72195"
FT   CDS             complement (50035..51666)
FT                   /*tag= an
FT                   /product= "ORF#31 protein"
FT                   /note= "encoded protein shown in AAW72196"
FT   CDS             complement (51701..53575)
FT                   /*tag= ao
FT                   /product= "ORF#32 protein"
FT                   /note= "encoded protein shown in AAW72197"
FT   CDS             54393..58115
FT                   /*tag= ap
FT                   /product= "ORF#33 protein"
FT                   /note= "encoded protein shown in AAW72198"
FT   CDS             complement (58060..58977)
FT                   /*tag= aq
FT                   /product= "ORF#34 protein"
FT                   /note= "encoded protein shown in AAW72199"
FT   CDS             complement (58970..60760)
FT                   /*tag= ar
FT                   /product= "ORF#35 protein"
FT                   /note= "encoded protein shown in AAW72200"
FT   CDS             60759..61151
FT                   /*tag= as
FT                   /product= "ORF#36 protein"
FT                   /note= "encoded protein shown in AAW72201"
```

```
FT   CDS             61241..62071
FT                   /*tag=  at

  Query Match             13.1%;  Score 64.4;  DB 19;  Length 117213;
  Best Local Similarity   55.9%;  Pred. No. 0.00048;
  Matches  143;  Conservative    0;  Mismatches  111;  Indels    2;  Gaps    1;

Qy      20 CGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCC 79
           | | ||| |   |   |||    | |  |  || |||||  |||||||  |      ||||
Db 113688 CCCTCCGCCTCCTCCTCCTCCGCCTCCTCCTCCTCCGCCTCCTCCTCCTCCGCCTCCTCC 113747

Qy      80 TCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCT 139
           || ||| ||||| ||  ||| ||| |||| ||| | ||  |  | |||||||   | ||
Db 113748 TCCTCCGCCTCCTCCTCCTCCGCCTCCTCCTCCTCCGCCTCCTCCTCCTCCGCCGCCGCT 113807

Qy     140 AACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTG 199
             | || |  |  ||  |||| ||  |   ||| || ||||   | || | |||| |  |
Db 113808 GGCGCCGGACCCTGCTGCCTCTGCGGC--TGCCCCCGCGCCGCGGGCGCCTGCGGCCCCG 113865

Qy     200 AGCGCGAGCGGGCGCCAGCCGATGAGTACGGTTAGGCCCTCGTTCGCCGCTGGGGCGGCC 259
             |||  |    || |  | |  |  | || |  |||| ||  |  ||| | |||||
Db 113866 CTCGCCGGGCACCGGCGCCAGCGGGCTCAGGCTCAGGCCCCGGGCCGCGCCGCGGCGGGA 113925

Qy     260 GGCCGGAGAGCTGCGG 275
           |  | | | |  | ||
Db 113926 GAACCGGGGGTGGGGG 113941


RESULT 5
ABQ52496/c
ID   ABQ52496 standard; DNA; 600 BP.
XX
AC   ABQ52496;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 39087.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 600 BP; 26 A; 51 C; 478 G; 44 T; 1 other;

  Query Match             12.0%;  Score 59;  DB 24;  Length 600;
  Best Local Similarity   58.1%;  Pred. No. 0.0035;
  Matches  104;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
           | ||   || |   | || ||||  |       ||   | |||  |  || || ||  |
Db     329 CGCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCC 270

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
           || ||||   || | || |  || || || || ||||  || || | |||  ||||  |
Db     269 CCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCAACTCCCCCCCCCCCCCCCCCCCCCG 210

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCC 181
           || ||||| |   |||   ||||| | | | || || || ||  ||| |||||  | |||
Db     209 TCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACGCCCCGCCCCCCC 151


RESULT 6
ABQ52497
ID   ABQ52497 standard; DNA; 600 BP.
XX
AC   ABQ52497;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 39088.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
```

CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 600 BP; 44 A; 478 C; 51 G; 26 T; 1 other;

Query Match 12.0%; Score 59; DB 24; Length 600;
Best Local Similarity 58.1%; Pred. No. 0.0035;
Matches 104; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
Db     272 CGCCCCCCGCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCC 331

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
Db     332 CCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCAACTCCCCCCCCCCCCCCCCCCCCCCCG 391

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCC 181
Db     392 TCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACGCCCCGCCCCCCCC 450
```

RESULT 7
AAS67812
ID AAS67812 standard; cDNA; 542 BP.
XX
AC AAS67812;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #3616.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG03625.
XX
PT New isolated polynucleotide and encoded polypeptides useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 3616; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 542 BP; 68 A; 211 C; 123 G; 139 T; 1 other;

Query Match 11.9%; Score 58.6; DB 23; Length 542;
Best Local Similarity 53.2%; Pred. No. 0.0042;
Matches 124; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```
Qy      44 CCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCC 103
Db     301 CCGTCCACCACCTCGACGACCACCTCCTCGTCCTCCTCTTCCTCGTCGTCATCCTCCTCC 360

Qy     104 ACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGC 163
Db     361 TCCGCGCCCCCCTCGGGGCTTCTCCTCCACACGCCCGAGCCCCCTCCGCCGCCTGCTCCCG 420

Qy     164 CTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATG 223
Db     421 CCCTCCTCGTCCTCGCCGTCCTCGTCGTCTTCGTCGCCGTCCACCACCTCGACGACCACC 480

Qy     224 AGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGC 276
Db     481 TCCTCGTCCTCCTCTTCCTCCTCTGCCTCCGCCGGTGGCCGCAGGGCCGGGAC 533
```

RESULT 8
AAX09011/c
ID AAX09011 standard; cDNA; 1266 BP.
XX
AC AAX09011;
XX
DT 14-JUN-1999 (first entry)
XX
DE Brn-3a polynucleotide.
XX
KW Brn-3a; Bcl-2; neurons; neuronal cells; apoptosis; cell death; CNS;
KW PNS; central nervous system; parasympathetic nervous system;
KW development; injury; neurotrophic factor; nerve growth factor; NGF;
KW ciliary neurotrophic factor; CNTF; brain-derived neurotrophic factor;
KW BNTF; neurotrophin; NT-3; NT-4; NT-5; neurodegenerative disease;
KW familial dysautonomia; infantile muscular dystrophy;
KW Parkinson's disease; Alzheimer's disease; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT /*tag= a
FT /product= Brn_3a_polypeptide
XX

PN WO9905272-A1.
XX
PD 04-FEB-1999.
XX
PF 27-JUL-1998; 98WO-GB02228.
XX
PR 10-DEC-1997; 97US-0988476.
PR 25-JUL-1997; 97GB-0015823.
XX
PA (UNLO) UNIV COLLEGE LONDON.
XX
PI Latchman DS, Smith MD;
XX
DR WPI; 1999-142928/12.
DR P-PSDB; AAW96263.
XX
PT New polypeptide comprising transcription factor Brn-3a, or its
PT derivative - useful for treating nervous system diseases, preventing
PT cellular apoptosis and increasing nerve regeneration following
PT neuronal damage
X
PS Disclosure; Page 63-64; 68pp; English.
XX
CC Over expression of transcription factor Brn-3a can protect neuronal
CC cells from apoptosis. Brn-3a also specifically activates expression
CC of the BCl-2 gene in neuronal cells and this activation is mediated
CC via a Brn-3a response element in the 5' regulatory region of the
CC Bcl-2 gene. Both the anti-apoptopic effect of Brn-3a and its ability
CC to activate expression of Bcl-2 are mediated by the N-terminal
CC domain of Brn-3a. Members of the Bcl-2 family perform critical roles
CC in the regulation of selective apoptosis during development of the
CC nervous system. The stimulation of Bcl-2 expression by Brn-3a in a
CC neuron specific manner and consequent protection of neuronal cells
CC from apoptosis suggests that Brn-3a may co-ordinate some aspects of
CC neuronal reorganisation during development or following injury. The
CC elevation of Brn-3a expression by either pharmacological means
CC (compositions comprising one other therapeutic polypeptide e.g.
CC neurotrophic factors, nerve growth factor (NGF), ciliary neurotrophic
CC factor (CNTF), brain-derived neurotrophic factor (BNTF), and
CC neurotrophins NT-3 and NT-4/5) or gene therapy may represent a
CC method for treating human diseases associated with excessive
CC neuronal cell death and/or lack of nerve regeneration, especially
CC neurodegenerative diseases such as familial dysautonomia and
CC infantile muscular dystrophy, and Parkinson's and Alzheimer's
CC disease.
XX
SQ Sequence 1266 BP; 201 A; 455 C; 458 G; 152 T; 0 other;

Query Match 11.8%; Score 58.2; DB 20; Length 1266;
Best Local Similarity 51.3%; Pred. No. 0.0055;
Matches 135; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy      40 GCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTC 99
           || | |  |    | ||||| | || ||||   |  | || |  || || |||||| |
Db     550 GCCCGCCGCCCGGGGCGCCGCCGCCGCCCCCCGGGCCGCCGCCGCCGCCGCCACCCCCGG 491

Qy     100 GGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCT 159
           |||| || |  ||    ||  |  |||| ||     | |   | ||| |     | || |
Db     490 GGCCGCCGCCGCCCGGGCCACCGCCGCCCCCCGGTCCGCCTCCGCCCCCGGGGCCGTCGT 431

Qy     160 CCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCC 219
             ||  |  |||||||||| | |    | | |||   ||    ||    || |||| |||
Db     430 GGGCGCCGCCGCCGCCTCCCGCCGCGCCTGCGCCCCCTGCGCCGGCCATGAGCGCGAGCG 371

Qy     220 GATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGT 279
           | |    | |  || |  |   | | | |||   |   |||  |   | | ||  ||
Db     370 ACGGCGACGAGATGTGGTCCAGCAGGTCACCGGGCTCGAGCGCCTGGTGGTGGTGGTGGT 311

Qy     280 GGGGGAGTTGGCGACGGAGCGGC 302
           || || | ||| |    ||||||
Db     310 GGTGGTGGTGGTGCGCCAGCGGC 288
```

RESULT 9
AAA29007/c
ID AAA29007 standard; cDNA; 1266 BP.
XX
AC AAA29007;
XX
DT 12-SEP-2000 (first entry)
XX
DE Murine transcription factor Brn-3a coding sequence.
XX
KW Brn-3a; modulator; inhibitor; cervical cancer; human papilloma virus;
KW HPV; antisense; cytostatic; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT /*tag= a
FT /product= Brn-3a_transcription_factor
XX
PN WO200034466-A1.
XX
PD 15-JUN-2000.
XX
PF 07-DEC-1999; 99WO-GB04116.
XX
PR 07-DEC-1998; 98GB-0026888.
PR 31-MAR-1999; 99US-0282210.
XX
PA (UNLO) UNIV COLLEGE LONDON.
XX
PI Latchman DS, Budhram-Mahadeo V, Ndisang D;
XX
DR WPI; 2000-423416/36.
DR P-PSDB; AAY96405.
XX
PT Product for treating, preventing and diagnosing cervical cancer
PT comprises a nucleotide sequence or molecule which binds to Brn-3a,
PT decreases its intracellular levels or inhibits its activity
XX
PS Disclosure; Page 65-68; 72pp; English.
XX
CC A product that binds, causes a decrease in intracellular levels of or
CC inhibits the activity of Brn-3a useful for treating, prevention or
CC diagnosis of cervical cancer caused by human papilloma virus (HPV) is
CC claimed. Expression of HPV proteins is generally dependent on the
CC presence of Brn-3a in the cell. Methods of identifying Brn-3a binding
CC agents or agents which inhibit Brn-3a expression are claimed. Nude mice
CC were injected with SiHa cells containing a single integrated HPV16-genome
CC were transformed with a Brn-3a antisense construct and with the empty
CC expression vector as control and tumours assessed at regular intervals.
CC Results showed that after 30 days there was no or very little tumour
CC growth in mice transformed with Brn-3a antisense construct as compared
CC to the control.
XX
SQ Sequence 1266 BP; 201 A; 455 C; 458 G; 152 T; 0 other;

Query Match 11.8%; Score 58.2; DB 21; Length 1266;
Best Local Similarity 51.3%; Pred. No. 0.0055;
Matches 135; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

```
Qy      40 GCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTC 99
           || | |  |    | ||||| | || ||||   |  | || |  || || |||||| |
Db     550 GCCCGCCGCCCGGGGCGCCGCCGCCGCCCCCCGGGCCGCCGCCGCCGCCGCCACCCCCGG 491

Qy     100 GGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCT 159
           |||| || |  ||    ||  |  |||| ||     | |   | ||| |     | || |
Db     490 GGCCGCCGCCGCCCGGGCCACCGCCGCCCCCCGGTCCGCCTCCGCCCCCGGGGCCGTCGT 431

Qy     160 CCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCC 219
             ||  |  |||||||||| | |    | | |||   ||    ||    || |||| |||
Db     430 GGGCGCCGCCGCCGCCTCCCGCCGCGCCTGCGCCCCCTGCGCCGGCCATGAGCGCGAGCG 371
```

```
Qy     220 GATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGT 279
           | |    | |  || |  |   | | | |||   |   ||| |   |  | ||  ||
Db     370 ACGGCGACGAGATGTGGTCCAGCAGGTCACCGGGCTCGAGCGCCTGGTGGTGGTGGTGGT 311

Qy     280 GGGGGAGTTGGCGACGGAGCGGC 302
           || || | ||| |    ||||||
Db     310 GGTGGTGGTGGTGCGCCAGCGGC 288


RESULT 10
ABK42449
ID   ABK42449 standard; DNA; 9953 BP.
XX
AC   ABK42449;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Genomic sequence #348 encoding novel human connective tissue polypeptide.
XX
"W   Human; connective tissue related disorder; cancer; gene therapy;
 I   cytostatic; gene; ds.
.X
OS   Homo sapiens.
XX
PN   WO200155343-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01322.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-02  063.
PR   14-SEP-2000; 2000US-02  064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
```

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-565190/63.
XX
PT Nucleic acid encoding novel connective tissue associated polypeptides,
PT used in diagnosing, preventing, treating or ameliorating a disorder
PT such as cancer or rheumatoid arthritis -
XX
PS Disclosure; SEQ ID No 1336; 673pp; English.
XX
CC The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful
CC in gene therapy. ABK42102-ABK43116 represent genomic sequences encoding
CC the novel human connective tissue related polypeptides.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9953 BP; 1958 A; 3052 C; 2783 G; 2160 T; 0 other;

Query Match 11.5%; Score 56.4; DB 23; Length 9953;
Best Local Similarity 58.2%; Pred. No. 0.016;
Matches 99; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy      21 GCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCT 80
           ||   ||  ||  |  ||| || ||  || ||||||||  || |||| |      |||||
Db    3991 GCTTTGAGCTTGGAGGCTTCCTGCCCGCAGCCGCCGCCTCCTTCTCCTCCTCCTCCTCCT 4050

Qy      81 CTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTA 140
           | ||||||||| ||  |||  || ||||  ||| | ||  |||| | |  |    |||
Db    4051 CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCCTCCTCCCTCCTCCTCCTC 4110

Qy     141 ACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCC 190
             | ||  ||| | ||||||||  |  |   |  |  | ||    ||  ||
Db    4111 CCTCCTCCTCCCTCCTCCTCCCTCCTCCTCCTCCTCCTCCTTCCCCC 4160
```

RESULT 11
ABK42448
ID ABK42448 standard; DNA; 11189 BP.
XX
AC ABK42448;
XX
DT 21-MAY-2002 (first entry)
XX
DE Genomic sequence #347 encoding novel human connective tissue polypeptide.
XX
KW Human; connective tissue related disorder; cancer; gene therapy;
KW cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200155343-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01322.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-565190/63.
XX
PT Nucleic acid encoding novel connective tissue associated polypeptides,
PT used in diagnosing, preventing, treating or ameliorating a disorder
PT such as cancer or rheumatoid arthritis -
XX
PS Disclosure; SEQ ID No 1335; 673pp; English.
XX
CC The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful
CC in gene therapy. ABK42102-ABK43116 represent genomic sequences encoding
CC the novel human connective tissue related polypeptides.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 11189 BP; 2196 A; 3373 C; 3188 G; 2432 T; 0 other;

Query Match 11.5%; Score 56.4; DB 23; Length 11189;
Best Local Similarity 58.2%; Pred. No. 0.016;
Matches 99 Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 21 GCCC JACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCT 80
Db 5227 GCTTTGAGCTTGGAGGCTTCCTGCCCGCAGCCGCCGCCTCCTTCTCCTCCTCCTCCTCCT 5286

Qy 81 CTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTA 140
Db 5287 CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCCTCCTCCCTCCTCCTCCTC 5346

Qy 141 ACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCC 190
Db 5347 CCTCCTCCTCCCTCCTCCTCCCTCCTCCTCCTCCTCCTCCTTCCCCC 5396

RESULT 12
AAI84343/c
ID AAI84343 standard; cDNA; 434 BP.
XX
AC AAI84343;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4403.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.

XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO04412.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 4403; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 434 BP; 34 A; 32 C; 256 G; 47 T; 65 other;

Query Match 11.4%; Score 56; DB 22; Length 434;
Best Local Similarity 41.8%; Pred. No. 0.014;
Matches 107; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```
Qy      17 CTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCCCCGACTCCTCCCCAATGCTC 76
           | || ||||  |       ||   | |||  |  || || ||  |  | ||||    | |
Db     387 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCC 328

Qy      77 TCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAAC 136
           || |  || || || ||| | |  || || | |||  |||| || | || || | |   |
Db     327 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 268

Qy     137 CCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTC 196
           ||   ||||| | | | || || || || || | || || | |||    |  |  |   |
Db     267 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCNCCNNCNNNC 208

Qy     197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCG 256
                                                             |
Db     207 CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNCA 148

Qy     257 GCCGGCCGGAGAGCTG 272
           ||  ||    |||||
Db     147 CCCCACCCCCCAGCTG 132
```

RESULT 13
ABI99656/c
ID ABI99656 standard; cDNA; 1286 BP.
XX
AC ABI99656;
XX
DT 07-MAR-2002 (first entry)
XX
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:686.
XX
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS Mus musculus.
XX
PN WO200188188-A2.
XX
PD 22-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-JP04192.
XX
PR 18-MAY-2000; 2000JP-0145977.
XX
PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX
DR WPI; 2002-034733/04.
XX
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
XX
PS Claim 2; Page 1713-1714; 2690pp; English.
XX
CC The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischaemic condition-improving
CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 1286 BP; 239 A; 252 C; 462 G; 173 T; 160 other;

Query Match 11.4%; Score 56; DB 24; Length 1286;
Best Local Similarity 32.4%; Pred. No. 0.016;
Matches 110; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```
Qy       3 CTCCGAGCGTCATTCTCCGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACT 62
           | ||   |  |   | || ||||  |       ||   | |||  |  || || ||  |
Db    1257 CNCCCCCCCCCCCCCCCCCCCCCNCCCCNCCCCCCCNCCCCCCCCCCCCCCCCCCCCC 1198

Qy      63 CCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCC 122
           || ||||    | | || |  || || || ||| | |  || || | ||   ||||  ||
Db    1197 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCNCCCCCCCCC 1138

Qy     123 TCGCCTCCGCAAACCCTAACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCG 182
            | || || |    ||   ||||| | | | || || || || || | | || |  | |||
Db    1137 CCCCCCCCCCNCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCC 1078

Qy     183 TGTCGTCCGCGGTCTTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGGTTAGGGCCTCGT 242
              |  || |   |     | |                    |
Db    1077 CCCCCCCCNCNCCCCCNCNCNCCNNCNCNNNNNNNNNNNNNNNNCNNNNNNNNNNNNN 1018

Qy     243 TCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGGGGAGTTGGCGACGGAGCGGC 302
                              |          |          |          |
Db    1017 NNNNNNNNNNNNNNNNNNCNNNNNNNNNNNNNCNNNNNNNNNNNNNNCNNNNNNNNNC 958

Qy     303 TGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGG 341
                                          | |
Db     957 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCNNNGTG 919
```

RESULT 14
AAS67813
ID AAS67813 standard; cDNA; 853 BP.
XX
AC AAS67813;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #3617.
XX

```
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG03626.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 3617; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 853 BP; 129 A; 320 C; 267 G; 137 T; 0 other;

  Query Match           11.3%;  Score 55.8;  DB 23;  Length 853;
  Best Local Similarity  51.2%;  Pred. No. 0.016;
  Matches  155;  Conservative    0;  Mismatches  147;  Indels    1;  Gaps    1;

Qy      44 CCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCCTCTTCCTCCTCCACCCACTCGGCC 103
           ||| | ||| || |     || || |   |  ||||||||||| ||  |   |||  ||
Db     444 CCGTCCACCACCTCGACGACCACCTCCTCGTCCTCCTCTTCCTCGTCGTCATCCTCCTCC 503

Qy     104 ACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCTAACC CCGCTCTCGCCTCCTCCGC 163
            || |||||   |  ||| || ||||| ||  ||| | ||||| |   |||||  ||||
Db     504 TCCGCGCCCCCTCGGGGCTTCTCCTCCACACGCCCGAGCCCCCTCCGCCGCCTGCTCCCG 563

Qy     164 CTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTGAGCGCGAGTG-GGCGCCAGCCGAT 222
           | |  | |  |||||  |    ||  |      |   |||  |  | |  | | |  || |
Db     564 CCCTCCTCGTCCTCGTCTTCCTCCCCTAGAAGCAGCAGCACCGGCGACGAGGCGGCGGCG 623

Qy     223 GAGTACGGTTAGGGCCTCGTTCGCCGCTGGGGCGGCCGGCCGGAGAGCTGCGGCAGTGGG 282
           | |  |    || ||  |   | |      |  | ||  |   || |||||||  ||
Db     624 GCGGCCCCGGTGGCCCGAGGGGGGCCGGTCCCGGCGCTGCGCGGCGGCGGCGGCAGCGGC 683

Qy     283 GGAGTTGGCGACGGAGCGGCTGGTGGAGGTGGCGCAACGGGCGGCGGACGCTGCTGGGGA 342
```

```
           ||  |      |   | ||| || |  || | ||   | ||    ||  || |  |
Db     684 GGCCTCCAGCTCTCCGGGGCGGGCGCCGGGGCCGGGCCCGGGACAGGGGGCGGGAGCCCG 743

Qy     343 GGT 345
           |||
Db     744 GGT 746


RESULT 15
ABI99807/c
ID   ABI99807 standard; cDNA; 1045 BP.
XX
AC   ABI99807;
XX
DT   07-MAR-2002  (first entry)
XX
DE   Mouse ischaemic condition related cDNA sequence SEQ ID NO:916.
XX
KW   Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW   vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
OS   Mus musculus.
XX
PN   WO200188188-A2.
XX
PD   22-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-JP04192.
XX
PR   18-MAY-2000; 2000JP-0145977.
XX
PA   (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI   Ishikawa K,  Asai S,  Takahashi Y,  Nagata T,  Ishii Y;
XX
DR   WPI; 2002-034733/04.
XX
PT   Examining the ischemic ondition (e.g. occlusiv ischemia) by measuring
PT   expression levels of particular genes defined in the specification or
PT   by determining the expression profile of a gene group comprising these
PT   genes -
XX
PS   Claim 2; Page 2313; 2690pp; English.
XX
CC   The present invention describes a method for examining ischaemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (I) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (I). The method
CC   is useful for examining the ischaemic condition (e.g. compressive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding
CC   the protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The
CC   expression levels or expression profiles produced by these genes are
CC   used as an indicator when screening for ischaemic condition-improving
CC   drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC   represent PCR primers for a mouse ischaemic condition related sequence,
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence 1045 BP; 228 A; 196 C; 398 G; 198 T; 25 other;

  Query Match           11.2%;  Score 55;  DB 24;  Length 1045;
  Best Local Similarity  49.5%;  Pred. No. 0.024;
  Matches  103;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy      20 CGCCCCGACATTTAAACCTTGCTCCCGACAACCGCCGCCGACTCCTCCCCAATGCTCTCC 79
           | ||||  |       ||   | |||  |  || || ||  | || ||||    | | ||
Db    1044 CCCCCCCCCCCCCCCCCCNCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 985

Qy      80 TCTTCCTCCTCCACCCACTCGGCCACCTCGCCCTTCCCCGGCCTCGCCTCCGCAAACCCT 139
           |  || || || ||| | |  || || | |||  ||||  || | || || |   |||
Db     984 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCC 925
```

```
Qy     140 AACCCCCGCTCTCGCCTCCTCCGCCTCCGCGCCGCCTCGCCCGTGTCGTCCGCGGTCTTG 199
           ||||| | | | || || || || || | |  | | |                ||| |
Db     924 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCNNCNNCNCCNNNNNNNNNNNNNNNNTCTCG 865

Qy     200 AGCGCGAGTGGGCGCCAGCCGATGAGTA 227
           || | |   ||   ||    | | ||
Db     864 TCCGTGGGCCCGCTGAAGTTTTTTAATA 837


Search completed: November 19, 2002, 00:26:51
Job time : 396 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 17, 2002, 14:30:07 ; Search time 26 Seconds
(without alignments)
539.192 Million cell updates/sec

Title: US-09-686-522C-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	334	19.3	261	1	YHEB_CHLVI	P56160 chlorobium
2	292	16.9	271	1	SUHB_PSEAE	Q9hxi4 pseudomonas
3	287	16.6	267	1	SUHB_PASMU	Q9cnv8 pasteurella
4	279	16.1	264	1	SUHB_AQUAE	O67791 aquifex aeo
5	272	15.7	267	1	SUHB_HAEIN	P44333 haemophilus
6	265.5	15.3	268	1	MYO3_LYCES	P54928 lycopersico
7	262	15.1	267	1	SUHB_SALTY	P58537 salmonella
8	258	14.9	267	1	SUHB_ECOLI	P22783 escherichia
9	248	14.3	267	1	SUHB_VIBCH	Q9kty5 vibrio chol
10	247.5	14.3	273	1	MYO1_LYCES	P54926 lycopersico
11	238	13.8	265	1	MYO2_LYCES	P54927 lycopersico
12	235	13.6	270	1	MYOP_MESCR	O49071 mesembryant
13	234	13.5	275	1	SUHB_XYLFA	Q9pam0 xylella fas
14	229.5	13.3	256	1	SUHB_THEMA	O33832 thermotoga
15	228	13.2	266	1	SUHB_RHILO	Q98f59 rhizobium l
16	221.5	12.8	265	1	SUHB_BACSU	Q45499 bacillus su
17	218	12.6	259	1	STRO_STRGR	P29785 streptomyce
18	212.5	12.3	287	1	SUHB_SYNY3	P74158 synechocyst
19	211	12.2	266	1	SUHB_RHIME	Q92m71 rhizobium m
20	211	12.2	277	1	MYOP_BOVIN	P20456 bos taurus
21	209	12.1	285	1	MYOP_XENLA	P29219 xenopus lae
22	206	11.9	246	1	CYSQ_ECOLI	P22255 escherichia
23	203.5	11.8	277	1	MYOP_MOUSE	O55023 mus musculu
24	202.5	11.7	261	1	SUHB_NEIMB	Q9jz07 neisseria m
25	201.5	11.6	261	1	SUHB_NEIMA	Q9ju03 neisseria m
26	200	11.6	252	1	SUHB_METJA	Q57573 methanococc
27	196.5	11.4	288	1	MYO2_HUMAN	O14732 homo sapien
28	193.5	11.2	277	1	MYOP_HUMAN	P29218 homo sapien
29	193.5	11.2	277	1	MYOP_RAT	P97697 rattus norv
30	190.5	11.0	269	1	CYSQ_ACTAC	P70714 actinobacil
31	190	11.0	246	1	CYSQ_SALTY	P26264 salmonella
32	183.5	10.6	256	1	SUHB_CAUCR	Q9a3d5 caulobacter
33	181.5	10.5	341	1	YSNM_CAEEL	Q19420 caenorhabdi
34	181	10.5	275	1	Y4FL_RHISN	P55450 rhizobium s
35	180.5	10.4	340	1	QAX_NEUCR	P11634 neurospora
36	175	10.1	280	1	SUHB_METTH	O26957 methanobact
37	171	9.9	353	1	DPNP_ARATH	Q42546 arabidopsis
38	164.5	9.5	252	1	SUHB_ARCFU	O30298 archaeoglob
39	161	9.3	266	1	SUHB_BUCAI	P57372 buchnera ap
40	158.5	9.2	358	1	DPNP_ORYSA	Q40639 oryza sativ
41	157	9.1	290	1	SUHB_MYCTU	O07203 mycobacteri
42	151.5	8.8	357	1	HAL2_YEAST	P32179 saccharomyc
43	144.5	8.4	265	1	CYSQ_BUCAI	P57624 buchnera ap
44	142	8.2	295	1	YHK6_YEAST	P38710 saccharomyc
45	141.5	8.2	330	1	QUTG_EMENI	P25416 emericella

ALIGNMENTS

RESULT 1
YHEB_CHLVI
ID YHEB_CHLVI STANDARD; PRT; 261 AA.
AC P56160;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hypothetical 28.2 kDa protein in hemB 3'region.
OS Chlorobium vibrioforme.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=f thiosulfatophilum / NCIB 8327;
RX MEDLINE=96215215; PubMed=8626508;
RA Rhie G.-E., Avissar Y.J., Beale S.I.;
RT "Structure and expression of the Chlorobium vibrioforme hemB gene and
RT characterization of its encoded enzyme, porphobilinogen synthase.";
RL J. Biol. Chem. 271:8176-8182(1996).
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U38348; -; NOT_ANNOTATED_CDS.
DR HSSP; P29218; 1IMF.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 261 AA; 28243 MW; D25E7D223024095F CRC64;

Query Match 19.3%; Score 334; DB 1; Length 261;
Best Local Similarity 33.9%; Pred. No. 9.9e-19;
Matches 78; Conservative 43; Mismatches 91; Indels 18; Gaps 4;

```
Qy      81 VAQRAADAAGEVLRKYF-RQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: ||::   || |: ::: | | :||| ||| ||| :   |   ||   :|||
Db       9 LALELAEKAGKLTLDYFGRRSLQVFSKRDDTPVTEADRNAEELIRQGISAKFPDDGLFGE 68

Qy     140 E-------NGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           |       || |         |::||||||:||| | ||:| :|||   |   :|||: |
Db      69 EFDEHPSGNGRR---------WIIDPIDGTRSFIHGVPLYGVMIALEVEGAMQLGVINFP 119

Qy     193 ILRERWIGVDGKQTTLNGQEISVRS-CNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKV 251
           | | :    |    :|| : | :     |   ::|   :| :  :     ::|    :
Db     120 ALGELYQAERGSGAFMNGSPVQVSAIAENSASTVVFTEKEYLLDPPSNHPVDQLRIDAGL 179

Qy     252 PLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
```

```
                 ||| : |:|||  :: |:  : |:|  :::|::| |||   |:||
Db     180 VRGWGDCYGHMLVASGRAEVAVDKIMSPWDCAAVIPIVEEAGGCCFDYRG 229


RESULT 2
SUHB_PSEAE
ID   SUHB_PSEAE     STANDARD;      PRT;   271 AA.
AC   Q9HXI4;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR PA3818.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004799; AAG07205.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   271 AA;  29521 MW;  EE1D43CFFEB871EA CRC64;

  Query Match             16.9%;  Score 292;  DB 1;  Length 271;
  Best Local Similarity   29.7%;  Pred. No. 1.8e-15;
  Matches   82;  Conservative   54;  Mismatches  106;  Indels   34;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEII---DKEDHSPVTIADREAEEAMVSVILKSFP 132
           :: :| ||| :|||::   ||  :|:::|   :|:    ||  || ||: :|: : |::|
Db       4 MLNIALRAARSAGELI---FRSIERLDVISVNEKDAKDYVTEVDRAAEQTIVAALRKAYP 60

Qy     133 THAIFGEENGW-RCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
           |||| ||| |:   :   ||::||:||:||| :|| | | |   ||  : |:    |:
Db      61 THAIMGEEGGFIEGSGEGADYLWVIDPLDGTTNFIHGVPHFAVSIACKYKGRLEHAVVLD 120

Qy     192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHL-FEADAEDAFIR-----V 245
           |: :| :   |:  |||: : |     | | | |   : |  | ::     |
Db     121 PVRQEEFTASRGRGAALNGRRLRVSGRKSLEGALLGTGFPFRDNQIDNLDNYLNMFRSLV 180

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG---- 301
           :   |      | :|:|  |   | ||  :|  .   ::: ||| ::|: |
Db     181 GQTAGIRRAGAASLDLAYVAAGRYDAFWEFGLSEWDMAAGALLVQEAGGLVSDFTGSHEF 240

Qy     302 -DKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                :| |              : ||: :  |  |  ::
Db     241 LEKGH-------------IVAGNTKCFKALLTTIQ 262


RESULT 3
SUHB_PASMU
ID   SUHB_PASMU     STANDARD;      PRT;   267 AA.
AC   Q9CNV8;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR PM0315.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006067; AAK02399.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29470 MW;  E7070FAE3BA589CA CRC64;

  Query Match             16.6%;  Score 287;  DB 1;  Length 267;
  Best Local Similarity   30.6%;  Pred. No. 4.2e-15;
  Matches   83;  Conservative   49;  Mismatches  111;  Indels   28;  Gaps    6;

Qy      78 LVEVAQRAADAAGEVLRKYFRQR--VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || |: | : :|  ::   |  :  ||  |: :|||:: || ||:| |
Db       4 MLNIAIRAARKAGNVIAKGYERRDDLQTTLKSTNDYVTNIDKASEEAIIEVIRKSYPDHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  ||:|    ::| |  ||:||:||| :|: | | |   ||:   |:  :||:  ||
Db      64 IITEESGALEGKDS-DIQWVIDPLDGTTNFVKGLPHFSVSIAIRVKGRTEVGVVYDPIRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           | :  | |:   :|   : | :   ||   | |  |           |  : | ::
Db     123 ELFTAVRGEGAKINDMRLRVENKRDLAGTVLTTGFPFKQTRLMPMQFAMMNNLIQ----- 177

Qy     256 CDCYAY----------ALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
            ||  :            :|:| ||   | |:| :|  :   ::  ||| : |: |
Db     178 -DCADFRRMGSAALDLCYVAAGRVDGYFEVGVKAWDIAAGDLIVREAGGLVCDFNG---- 232

Qy     306 WPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                 |  || ::|||   |: || |: ::
Db     233 ----GHSYLTSGHLVAAA-PRIVKEILNKIQ 258
```

```
RESULT 4
SUHB_AQUAE
ID   SUHB_AQUAE     STANDARD;      PRT;   264 AA.
AC   O67791;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR IMP2 OR AQ_1983.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC   Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000766; AAC07753.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; FALSE_NEG.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   264 AA;  29336 MW;  10F32D1D3B81B76A CRC64;

  Query Match             16.1%;  Score 279;  DB 1;  Length 264;
  Best Local Similarity   28.5%;  Pred. No. 1.7e-14;
  Matches   75;  Conservative   54;  Mismatches  106;  Indels   28;  Gaps    7;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  |||| ||
Db       5 KKYLEVAKIAALAGGQVLKENFGKVKKENIEEKGEKDFVSYVDKTSEERIKEVILKFFPD 64

Qy     134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           | : ||| |   || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 HEVVGEEMG---AEGSGSEYRWFIDPLDGTKNYINGFPIFAVSVGLVKGEEPIVGAVYLP 121

Qy     193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTT.PHLFEADA-------LDAFIRV 245
           : : |  |    :||: | |:    |  | :    |      |           :| |  |
Db     122 YFDKLYWGAKGLGAYVNGKRIKVKDNESLKHAGVVYGFPSRSRRDISIYLNIFKDVFYEV 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
           :: |  |       ::| |  | ::|  :||:|  : : ::: |||
Db     182 -GSMRRP--GAAAVDLCMVAEGIFDGMMEFEMKPWDITAGLVILKEAGG---------- 227

Qy     306 WPVTAESRPTSFNVVAAGDARVH 328
           |    |  : : ||:  :|
Db     228 -VYTLVGEPFGVSDIIAGNKALH 249


RESULT 5
SUHB_HAEIN
ID   SUHB_HAEIN     STANDARD;      PRT;   267 AA.
AC   P44333;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR HI0937.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32775; AAC22595.1; -.
DR   HSSP; P29218; 1IMF.
DR   TIGR; HI0937; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29499 MW;  DF974E0DAB0CCEA6 CRC64;

  Query Match             15.7%;  Score 272;  DB 1;  Length 267;
  Best Local Similarity   30.8%;  Pred. No. 6e-14;
  Matches   84;  Conservative   44;  Mismatches  113;  Indels   32;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFRQR--VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || |: | : :|  :|   |  :  ||  |: :|  :: || ||:| |
Db       4 MLNIAIRAARKAGNVIAKNYERRDAIESTQKGINDYVTNVDKASEAEIIEVIRKSYPDHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |       :|  |::||:|||::|:|| | |   ||:    :  :||:  ||
Db      64 IITEETG-AIEGKDSDVQWIIDPLDGTRNFMTGLPHFSVSIAVRVKNRTEVGVVYDPIRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPH----------LFEADAEDAF-I 243
           | :  | |:   ||   : | |   :  : | |  |           :  |  |||
Db     123 ELFTAVRGEGAKLNEVRLRVDSKREIQGSILATGFPFKQPKLMPAQFAMMNALIEDAADF 182

Qy     244 RVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           |      : |    ||    :||  :|   | ||| :|  :   ::  ||| : |:
Db     183 RRTGSAALDL----CY----VASNRIDGYFEMGLKAWDCAAGDLIVREAGGLVCDFD--- 231

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                 |   | |::||  :|| || |: :|
```

```
Db    232 -----AGNSYLRSGNIIAA-PSRVIKEMLNKIR 258


RESULT 6
MYO3_LYCES
ID   MYO3_LYCES     STANDARD;      PRT;   268 AA.
AC   P54928;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase 3 (EC 3.1.3.25) (IMPase 3) (IMP
DE   3) (Inositol monophosphatase 3).
GN   IMP3.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. VFNT Cherry;
RX   MEDLINE=96351935; PubMed=8718627;
RA   Gillaspy G.E., Keddie J.S., Oda K., Gruissem W.;
RT   "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RT   by a multigene family.";
RL   Plant Cell 7:2175-2185(1995).
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39059; AAB19031.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Lithium; Multigene family.
SQ   SEQUENCE   268 AA;  29054 MW;  BCB635A029804901 CRC64;

  Query Match             15.3%;  Score 265.5;  DB 1;  Length 268;
  Best Local Similarity   30.2%;  Pred. No. 1.9e-13;
  Matches   83;  Conservative   48;  Mismatches  117;  Indels   27;  Gaps    8;

Qy     74 ATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
          : |: ::||  ||  |||::|: | :   :  |     ||  |:  |: : : : : ||:
Db      6 SVEQFLDVAVEAAKKAGEIIREGFYKTKHVEHKGMVDLVTETDKACEDFIFNHLKQRFPS 65

Qy    134 HAIFGEENGWRCA--ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
          |   |||    |   | : :  |::|||:||| :|: | |     | |    || :||:
Db     66 HKFIGEETTAACGNFELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIEKKPTVGVVYN 125

Qy    192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT---TSPHLFEADA-----EDAFI 243
          ||: | : |:|||   |||: | | | : | :| | |   |:      ||
Db    126 PIIDELFTGIDGKGAFLNGKPIKVSSQSELVKALLATEAGTNRDKLVVDATTGRINSLLF 185

Qy    244 RVRNKVKVPLYGCDCYAYAL--LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
          :||:     |  |   |  |  :| | :|:  |     |:|      ::: ||| : |
Db    186 KVRS-----LRMCGSCALNLCGVACGRLDLFYELEFGGPWDVAGGAVIVKEAGGFVFDPS 240
```

```
Qy    301 GDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
          | :  : :||         |||  :| :    : ||
Db    241 GSE--FDLTARR-------VAATNAHLKDAFIKAL 266


RESULT 7
SUHB_SALTY
ID   SUHB_SALTY     STANDARD;      PRT;   267 AA.
AC   P58537;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR STM2546.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008815; AAL21440.1; -.
DR   StyGene; SG?????; suhB.
DR   InterPro; IPR000760; Inositol_P.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29158 MW;  FC46D8A8298BB334 CRC64;

  Query Match             15.1%;  Score 262;  DB 1;  Length 267;
  Best Local Similarity   30.0%;  Pred. No. 3.5e-13;
  Matches   80;  Conservative   44;  Mismatches  123;  Indels   20;  Gaps    7;

Qy     78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
          :: :| |||  || |: | :     ||   |  :  ||  |: ||  ::  | ||:| |
Db      4 MLTIAVRAARKAGNVIAKNYETPDAVEASQKGSNDFVTNVDKAAEAVIIDTIRKSYPQHT 63

Qy    136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
          |  ||:|     :  |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db     64 IITEESGEHVGTDQ-DVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy    196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKVKVPL- 253
          | :    |:   ||| :   :   |    | |  |  |:|      :| :  |:
Db    123 ELFTATRGQGAQLNGYRLRGSTARDLDGTILATGFP--FKAKQYATTYINIIGKLFTECA 180

Qy    254 ----YGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
              |      | :|:| ||   | ||:|:|| :   ::  ||| ::|: |   :: :|
Db    181 DFRRTGSAALDLAYVAAGRVDGFFEIGLRPWDFAAGELLVREAGGIVSDFTGGH-NYMMT 239

Qy    310 AESRPTSFNVVAAGDARVHKEALDALR 336
```

```
                    : ||: || |  |  :|
Db    240 GN--------IVAGNPRVVKAMLANMR 258


RESULT 8
SUHB_ECOLI
ID   SUHB_ECOLI     STANDARD;      PRT;   267 AA.
AC   P22783; P77511; Q8X2E6;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR SSYA OR B2533 OR Z3800 OR ECS3399.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=90202735; PubMed=2138605;
RA   Yano R., Nagai H., Shiba K., Yura T.;
RT   "A mutation that enhances synthesis of sigma 32 and suppresses
RT   temperature-sensitive growth of the rpoH15 mutant of Escherichia
RT   coli.";
RL   J. Bacteriol. 172:2124-2130(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrha ic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [6]
RP   FUNCTION.
RX   MEDLINE=95095968; PubMed=8002619;
RA   Matsuhisa A., Suzuki N., Noda T., Shiba K.;
RT   "Inositol monophosphatase activity from the Escherichia coli suhB gene
RT   product.";
RL   J. Bacteriol. 177:200-205(1995).
RN   [7]
RP   CHARACTERIZATION.
RX   MEDLINE=20213266; PubMed=10747806;
RA   Chen L., Roberts M.F.;
RT   "Overexpression, purification, and analysis of complementation
RT   behavior of E. coli SuhB protein: comparison with bacterial and
RT   archaeal inositol monophosphatases.";
RL   Biochemistry 39:4145-4153(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M34828; AAA67506.1; -.
DR   EMBL; AE000339; AAC75586.1; -.
DR   EMBL; D90883; BAA16427.1; -.
DR   EMBL; D90884; BAA16435.1; -.
DR   EMBL; AE005483; -; NOT_ANNOTATED_CDS.
DR   EMBL; AP002562; BAB36822.1; -.
DR   PIR; A35158; A35158.
DR   HSSP; P29218; 1IMF.
DR   SWISS-2DPAGE; P22783; COLI.
DR   EcoGene; EG10983; suhB.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
FT   CONFLICT    141    141       R -> L (IN REF. 1).
SQ   SEQUENCE   267 AA;  29172 MW;  8FEC3508BD111301 CRC64;

  Query Match             14.9%;  Score 258;  DB 1;  Length 267;
  Best Local Similarity   29.6%;  Pred. No. 7.2e-13;
  Matches   79;  Conservative   44;  Mismatches  124;  Indels   20;  Gaps    7;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || :: | :     ||   |  :  ||  |: ||  ::  | ||:| |
Db       4 MLNIAVRAARKAGNLIAKNYETPDAVEASQKGSNDFVTNVDKAAEAVIIDTIRKSYPQHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  ||:|         |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db      64 IITEESG-ELEGTDQDVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKV----- 249
           | :    |:  |||  :   :   |    | |  |  |:|      :| :  |:
Db     123 ELFTATRGQGAQLNGYRLRGSTARDLDGTILATGFP--FKAKQYATTYINIVGKLFNECA 180

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
                 |      | :|:| ||   | ||:|:|| :   ::  ||| ::|: |   :: :|
Db     181 DFRRTGSAALDLAYVAAGRVDGFFEIGLRPWDFAAGELLVREAGGIVSDFTGGH-NYMLT 239

Qy     310 AESRPTSFNVVAAGDARVHKEALDALR 336
                   : ||: || |  |  :|
Db     240 GN--------IVAGNPRVVKAMLANMR 258
```

```
RESULT 9
SUHB_VIBCH
ID   SUHB_VIBCH     STANDARD;      PRT;   267 AA.
AC   Q9KTY5;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   VC0745.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004160; AAF93910.1; ALT_INIT.
DR   HSSP; P29218; 1IMF.
DR   TIGR; VC0745; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   267 AA;  29088 MW;  88AD8CDF78A7AD87 CRC64;

  Query Match             14.3%;  Score 248;  DB 1;  Length 267;
  Best Local Similarity   28.9%;  Pred. No. 4.2e-12;
  Matches   79;  Conservative   41;  Mismatches  121;  Indels   32;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  ||  : |     ::::   |  :  ||  |:|||  :|| |  |:| |
Db       4 MLNIAIRAARKAGNHIAKSLENAEKIQTTQKGSNDFVTNVDKEAEAIIVSTIKSSYPEHC 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |       :  |::||:||| :|: | | |   ||:   ||  :  :  |:
Db      64 IIAEEGG-LIEGKDKEVQWIIDPLDGTTNFVKGFPHFAVSIAVRFRGKTEVACVYDPMTN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           | :    |    ||   | |:    |  | | |  |   :  :| :|::: : : |
Db     123 ELFTAQRGAGAQLNNARIRVQPIKDLQGAVLATAFPFKQKQHSE-SFMKILSAMFVE--- 178

Qy     256 CD-----------CYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           |             ||    ||:  ||   | ||||:|  :   :   ||  :||: |
Db     179 CADFRRTGSAALDLCY----LAANRVDGYFELGLKPWDMAAGELIAREAGAIVTDFAG-- 232

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                     | |:||:    | |  |  :|
```

```
Db     233 ------GTDYMQSGNIVASSPRGV-KAILQHIR 258


RESULT 10
MYO1_LYCES
ID   MYO1_LYCES     STANDARD;      PRT;   273 AA.
AC   P54926;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase 1 (EC 3.1.3.25) (IMPase 1) (IMP
DE   1) (Inositol monophosphatase 1).
GN   IMP1.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. VFNT Cherry;
RX   MEDLINE=96351935; PubMed=8718627;
RA   Gillaspy G.E., Keddie J.S., Oda K., Gruissem W.;
RT   "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RT   by a multigene family.";
RL   Plant Cell 7:2175-2185(1995).
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39444; AAB19030.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Lithium; Multigene family.
SQ   SEQUENCE   273 AA;  29558 MW;  92D1D532E359E0FD CRC64;

  Query Match             14.3%;  Score 247.5;  DB 1;  Length 273;
  Best Local Similarity   28.9%;  Pred. No. 4.8e-12;
  Matches   80;  Conservative   53;  Mismatches  111;  Indels   33;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           |  : ||  ||  |||::|| | :   :: |     ||  |:  |: : : : : ||:|
Db       8 EEFLGVAVDAAKRAGEIIRKGFHETKHVVHKGQVDLVTETDKACEDLIFNHLKQHFPSHK 67

Qy     136 IFGEENGWRCAENSADF------VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           |||     :  : ||       |::||:||| :|: | |     | |     | :||:
Db      68 FIGEET----SAATGDFDLTDEPTWIVDPVDGTTNFVHGFPSVCVSIGLTIGKIPTVGVV 123

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYL----YTTSPHLFEADAEDAFIRV 245
           ||: | : |::||   |||: | | | : | :: |     ||  :|     |    |:
Db     124 YDPIIDELFTGINGKGAYLNGKPIKVSSQSELVKSLLGTEVGTTRDNL---TVETTTRRI 180

Qy     246 RN---KVKVPLYGCDCYAYAL--LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDW 299
           |   ||:  |  |   |  | :| | :::    |   |:|      ::: ||| : |
Db     181 NNLLPKVR-SLRMCGSCALDLCWVACGRLELFYLIGYGGPWDVAGGAVIVKEAGGVLFDP 239
```

```
Qy     300 RGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           | :  : :|::        ||| :  : :  ::||:
Db     240 SGSE--FDITSQR-------VAATNPHLKEAFVEALQ 267


RESULT 11
MYO2_LYCES
ID   MYO2_LYCES     STANDARD;      PRT;   265 AA.
AC   P54927;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase 2 (EC 3.1.3.25) (IMPase 2) (IMP
DE   2) (Inositol monophosphatase 2).
GN   IMP2.
OS   Lycopersicon esculentum (Tomato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4081;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. VFNT Cherry;
RX   MEDLINE=96351935; PubMed=8718627;
RA   Gillaspy G.E., Keddie J.S., Oda K., Gruissem W.;
RT   "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RT   by a multigene family.";
RL   Plant Cell 7:2175-2185(1995).
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39443; AAB19029.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Lithium; Multigene family.
SQ   SEQUENCE   265 AA;  28705 MW;  5369735F30FC884B CRC64;

  Query Match             13.8%;  Score 238;  DB 1;  Length 265;
  Best Local Similarity   30.8%;  Pred. No. 2.5e-11;
  Matches   84;  Conservative   46;  Mismatches  117;  Indels   26;  Gaps    9;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           |  ||:||  ||  |||::|  | :   |  |      ||  |:  |  : : : : ||:|
Db       2 EEFVDVAIEAAKKAGEIIRHGFYKSKHIEHKGVVDLVTETDKACEVLIFNHLKQCFPSHK 61

Qy     136 IFGEENGWRCAENSADF------VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           |||     |  | :|       |::||:||| :|: | |     | |    |||:||:
Db      62 FIGEET---TAAASGNFELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIEKKPVVGVV 118

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT---TSPHLFEADAEDAFI-RV 245
           ||: | :  : |:   |||: | | | : | :| : |   |:      ||    | ||
Db     119 YNPIIDELFTAIYGRGAFLNGKSIRVSSESQLVKALVATEVGTNRDKAIVDATTGRINRV 178

Qy     246 RNKVK-VPLYGCDCYAYALLASGFVDIV--VESGLKPYDFLSLVPVIEGAGGSITDWRGD 302
           ||: : : |        :| | :|:   :| |  |:|  :   ::  ||| : |  |
Db     179 IFKVRSLRMSGSCALNLCGVACGRLDLFYEIEFG-GPWDVAAGALIVIEAGGLVLDPSGS 237

Qy     303 KLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           :  : :||          ||| :| :    ::||
Db     238 E--FDLTARR-------VAATNAHLKDAFINAL 261


RESULT 12
MYOP_MESCR
ID   MYOP_MESCR     STANDARD;      PRT;   270 AA.
AC   O49071;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myo-inositol-1(or 4)-monophosphatase (EC 3.1.3.25) (IMPase) (IMP)
DE   (Inositol monophosphatase).
GN   IMP1.
OS   Mesembryanthemum crystallinum (Common ice plant).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
OX   NCBI_TaxID=3544;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nelson D.E., Bohnert H.J.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC       FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC   -!- ENZYME REGULATION: INHIBITED BY LI(+) (BY SIMILARITY).
CC   -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC       PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF037220; AAB92418.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Hydrolase; Magnesium; Lithium.
SQ   SEQUENCE   270 AA;  29161 MW;  0C045EB33D903A6C CRC64;

  Query Match             13.6%;  Score 235;  DB 1;  Length 270;
  Best Local Similarity   29.0%;  Pred. No. 4.3e-11;
  Matches   72;  Conservative   35;  Mismatches  115;  Indels   26;  Gaps    5;

Qy      85 AADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGEEN--G 142
           ||  ||||:|| |  :  :  |     ||  |:  |: : : : : :| |   |||
Db      17 AAKRAGEVIRKGFYVKKNVEHKGQVDLVTETDKSCEDIIFNCLKQQYPNHKFIGEETTAA 76

Qy     143 WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVD 202
           :   | : :  |::||:||| :|: | |     | |     | :||:  ||: | : ||
Db      77 YGATELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIGKVPTVGVVYNPIMNELFTGVR 136

Qy     203 GKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDA------------FIRVRNKV 249
           :   |||  | | | : |    | |      :    ||              ||:
Db     137 RQGAFLNGVPIHVSSKDELVNCLLVTEVGTKRDKSTVDATTNRINGLLFKVRSIRMAGSC 196

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWRGDKLHWPV 308
           : | |  |        |  |:: |:|    :|  : : ::| ||| | |  |      |
Db     197 ALDLCGIAC--------GRADLMYENGYGGAWDVTAGIVIVEEAGGVIFDPSGKDFDITV 248

Qy     309 T--AESRP 314
```

```
          |  | | |
Db    249 TRIAASNP 256


RESULT 13
SUHB_XYLFA
ID   SUHB_XYLFA     STANDARD;      PRT;   275 AA.
AC   Q9PAM0;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR XF2476.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004055; AAF85274.1; -.
DR   HSSP; P29218; 2HHM.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   275 AA;  30233 MW;  74F622ACC4EE082C CRC64;

  Query Match             13.5%;  Score 234;  DB 1;  Length 275;
```

```
  Best Local Similarity   28.4%;  Pred. No. 5.2e-11;
  Matches   73;  Conservative   51;  Mismatches  117;  Indels   16;  Gaps    8;

Qy      79 VEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAI 136
           | :  :|| :|| || ::    : : :| |       :  |  ||: :|  : :::| : |
Db       6 VNIMVKAARSAGNVLLRHINKLETLHVIQKSRMDYASDVDEMAEKVIVKELKRAYPEYGI 65

Qy     137 FGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
            ||| |    : :   :||:|||:||| ::: | | :   |||: ||:|   ||  |:  |
Db      66 LGEEGG---LQGNHRIMWVIDPLDGTSNYLRGFPHYCISIALVENGEPTDAVIFDPLRNE 122

Qy     197 RWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT-TSP-HLFEADAE----DAFIRVRNKVK 250
           :    |    || ::| | :   |    | |  ||     | |:    || :     ::
Db     123 LFTASRGAGAILNERKIRVANRKDLNGTMLNTGFSPRERSRAHAQLKCVDALLMQAEDIR 182

Qy     251 VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
           |      | :| |  |   |:|:| :|  : : ::  ||| : |::|     |
Db     183 --RSGSAALDLAYVACGRADAYFEAGIKVWDVAAGMLLVREAGGYVCDFKGADA--PRMD 238

Qy     311 ESRPTSFNVVAAGDARV 327
           :  | |  :| ||: :|
Db     239 DKGPESCQLV-AGNIKV 254


RESULT 14
SUHB_THEMA
ID   SUHB_THEMA     STANDARD;      PRT;   256 AA.
AC   O33832;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR TM1415.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC   Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=98386670; PubMed=9720201;
RA   Liebl W., Brem D., Gotschlich A.;
RT   "Analysis of the gene for beta-fructosidase (invertase, inulinase) of
RT   the hyperthermophilic bacterium Thermotoga maritima, and
RT   characterisation of the enzyme expressed in Escherichia coli.";
RL   Appl. Microbiol. Biotechnol. 50:55-64(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=99437848; PubMed=10508089;
RA   Chen L., Roberts M.F.;
RT   "Characterization of a tetrameric inositol monophosphatase from the
RT   hyperthermophilic bacterium Thermotoga maritima.";
RL   Appl. Environ. Microbiol. 65:4559-4567(1999).
CC   -!- FUNCTION: Displays a 20-fold higher rate of hydrolysis of the D
CC       isoform of inositol 1-phosphate than of the L isoform.
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium.
CC   -!- SUBUNIT: Homotetramer.
```

```
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 22.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ001073; CAA04517.1; -.
DR   EMBL; AE001794; AAD36486.1; ALT_FRAME.
DR   HSSP; P29218; 1IMF.
DR   TIGR; TM1415; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   256 AA;  28647 MW;  550C15FEEA50B8DC CRC64;

  Query Match             13.3%;  Score 229.5;  DB 1;  Length 256;
  Best Local Similarity   28.7%;  Pred. No. 1.1e-10;
  Matches   68;  Conservative   44;  Mismatches  104;  Indels   21;  Gaps    7;

Qy     100 RVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEENGWRCAENSADFVWVLD 157
           ||: ::|:      ||  ||||:  :|  | | ||   |  ||  :       | :|::|
Db      24 RVDNVEKKTGFKDIVTEIDREAQRMIVDEIRKFFPDENIMAEEGIF----EKGDRLWIID 79

Qy     158 PIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQEISVRS 217
           ||||| :|: | | |   :| : ||:  :||:  | | |     :|     ||: | | |
Db      80 PIDGTINFVHGLPNFSISLAYVENGEVKLGVVHAPALNETLYAEEGSGAFFNGERIRV-S 138

Qy     218 CNLLAQAYLYTTSPHLFEADAEDAFIRVRNK--VKVPLYGCDCYAYALLASGFVDIVVES 275
             |   :  : :|  ::   |    ||    |   :: : |      | : :| ||  |
Db     139 ENASLEECVGSTGSYV---DFTGKFIERMEKRTRRIRILGSAALNAAYVGAGRVDFFVTW 195

Qy     276 GLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEAL 332
           : |:|  : : ::: ||| :||: |         |:   | | : : :  :| | :
Db     196 RINPWDIAAGLIIVKEAGGMVTDFSG--------KEANAFSKNFIFS-NGLIHDEVV 243


RESULT 15
SUHB_RHILO
ID   SUHB_RHILO     STANDARD;      PRT;   266 AA.
AC   Q98F59;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE   phosphatase) (I-1-Pase).
GN   SUHB OR MLR3921.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
CC   -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC       inositol + phosphate.
CC   -!- COFACTOR: Magnesium (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP003003; BAB50708.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hydrolase; Magnesium; Complete proteome.
SQ   SEQUENCE   266 AA;  28655 MW;  843436E21ED326CF CRC64;

  Query Match             13.2%;  Score 228;  DB 1;  Length 266;
  Best Local Similarity   28.3%;  Pred. No. 1.5e-10;
  Matches   77;  Conservative   47;  Mismatches  128;  Indels   20;  Gaps    5;

Qy      73 LATERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKS 130
           :|   |: |  :||  ||  | : |   | ::: |     |: |||:||: : : : |:
Db       1 MARSALLNVMVQAAMKAGRSLSRDFGEVQNLQVSLKGPGDYVSQADRKAEDIIFAELSKA 60

Qy     131 FPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVID 190
             | :    || |    |:|    |::||:||| :|: | |||   |||   |: | |||
Db      61 RPGYGFLMEERGAVEGEDS-QHRWIVDPLDGTTNFLHGIPLFAVSIALERQGQIVAGVIY 119

Qy     191 QPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVK 250
             | : | :    |    :| : : |     |    :    |||      :  : :|| :
Db     120 NPAMDELYTTERGGGAFMNDRRLRVAGRIKLVDTVIGCGMPHLGRGHHGNFLVELRNVMA 179

Qy     251 ----VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG--DKL 304
                 |   |       | :|:| :|   |:||  :|  : : :|  ||| ::|  |  | |
Db     180 EVSGVRRLGSAALDLAYVAAGRMDGFWETGLSAWDIAAGLLLIREAGGFVSDMDGGQDML 239

Qy     305 HWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                             | ||:  : :  | |::
Db     240 -----------DNGSVVAGNEVIQRALLKAVK 260


Search completed: November 17, 2002, 16:39:37
Job time : 30 secs
```

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AG3034
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44693.1; PID:g17742322; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3885
A;Map position: linear chromosome
C;Superfamily: suppressor protein suhB

Query Match 30.3%; Score 525; DB 2; Length 262;
Best Local Similarity 42.7%; Pred. No. 7.3e-34;
Matches 112; Conservative 51; Mismatches 73; Indels 26; Gaps 7;

```
Qy      84 RAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
           : ||||       ||  : :|:|:|  : |||  |: || |: ::| : || | | |||:
Db      14 KLADAASAETLPRFRTGIAVINKQDGGYDPVTEGDQAAETAIRALIEERFPQHGILGEEH 73

Qy     142 GWRCAEN---SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
           |     |     | :||:||||||::||:| |::||||    :|:  :|::|||  :||:
Db      74 G-----NVGLDRDHIWVIDPIDGTRAFISGVPVWGTLIGFQSSGRATMGIMDQPFTKERY 128

Qy     199 IGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
           |||     |    ::|  | |  |: | |:||:||:| |: :  : :|:::|::  |
Db     129 F-ADGKAAWYFGPDGEKKIRTRDCASLSDAVLFTTTPHIFTAEEKPLYEKVQDQVRLFRY 187

Qy     255 GCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRP 314
           | ||||| |||:| ||:|:|||||||  :|:|||| |||::| | |           ||
Db     188 GVDCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAGGTMTTWDG----------GRP 237

Qy     315 -TSFNVVAAGDARVHKEALDAL 335
            ::|||   ||:|||  |
Db     238 ENGGRILAAGSKAVHEEALAIL 259
```

RESULT 3
E98251
similar to mlcb1779.29 (probable monophosphatase gene) in mycobacterium leprae [importe
Species: Agrobacterium tumefaciens
Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 15-Feb-2002
C;Accession: E98251
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: E98251
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-297 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89535.1; PID:g15159414; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1926
A;Map position: linear chromosome
C;Superfamily: suppressor protein suhB

Query Match 30.3%; Score 525; DB 2; Length 297;
Best Local Similarity 42.7%; Pred. No. 8.6e-34;
Matches 112; Conservative 51; Mismatches 73; Indels 26; Gaps 7;

```
Qy      84 RAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
           : ||||       ||  : :|:|:|  : |||  |: || |: ::| : || | | |||:
Db      49 KLADAASAETLPRFRTGIAVINKQDGGYDPVTEGDQAAETAIRALIEERFPQHGILGEEH 108

Qy     142 GWRCAEN---SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
           |     |     | :||:||||||::||:| |::||||    :|:  :|::|||  :||:
Db     109 G-----NVGLDRDHIWVIDPIDGTRAFISGVPVWGTLIGFQSSGRATMGIMDQPFTKERY 163

Qy     199 IGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
           |||     |    ::|  | |  |: | |:||:||:| |: :  : :|:::|::  |
Db     164 F-ADGKAAWYFGPDGEKKIRTRDCASLSDAVLFTTTPHIFTAEEKPLYEKVQDQVRLFRY 222

Qy     255 GCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRP 314
           | ||||| |||:| ||:|:|||||||  :|:|||| |||::| | |           ||
Db     223 GVDCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAGGTMTTWDG----------GRP 272

Qy     315 -TSFNVVAAGDARVHKEALDAL 335
            ::|||   ||:|||  |
Db     273 ENGGRILAAGSKAVHEEALAIL 294
```

RESULT 4
F95293
probable inositol monophosphatase [imported] - Sinorhizobium meliloti (strain 1021) m
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: F95293
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; E
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: F95293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-608 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK64912.1; PID:g14523333; GSPDB:GN00165
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMa0483
A;Genome: plasmid

Query Match 27.9%; Score 483; DB 2; Length 308;
Best Local Similarity 39.6%; Pred. No. 1.8e-30;
Matches 107; Conservative 38; Mismatches 103; Indels 22; Gaps 4;

```
Qy      70 VGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILK 129
           :|| |:      |   || | : :      | | | | | |||  ||  |: :   |
Db      56 LGEFAS-----FAHDIADLARQTISSAAGVRREPIAKSDASPVTETDRAVEKCLRERIAD 110

Qy     130 SFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           || | : ||| |     |  :||||:|||||||:|: | |::||||:|   | |::|:|
Db     111 HFPDHGVLGEEFGAEGLGN--EFVWVIDPIDGTKAFVAGLPVYGTLISLTRGGTPILGLI 168

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           | |:  :||:|| |: ||||   |   |   || |::   :|  |    :     :|
Db     169 DNPMTGDRWLGVSGQPTTLNNVPIRTASTTALATAFIANGNPDAFSPADKSRVESLRRIT 228

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
           :  :||  | ||  :| | ||| :: || |||: :|||||  |||| ||||:|
Db     229 RWCVYGGSCIAYGRVADGSVDISIDGGLDPYDYCALVPVITGAGGCITDWQG-------- 280

Qy     310 AESRPTSFN----VVAAGDARVHKEALDAL 335
           || : |     ||     :|:  |: |
Db     281 ---RPLTLNSGGLCVATATDLLHRHVLEIL 307
```

```
RESULT   5
C87528
inositol monophosphatase family protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: C87528
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87528
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-278 <STO>
A;Cross-references: GB:AE005673; NID:g13423761; PIDN:AAK24223.1; GSPDB:GN00148
  Genetics:
C;Gene: CC2252

  Query Match             27.6%;  Score 477;  DB 2;  Length 278;
  Best Local Similarity   40.8%;  Pred. No. 4.7e-30;
  Matches  116;  Conservative   48;  Mismatches   88;  Indels   32;  Gaps    9;

Qy      73 LATERLVEVAQRAAD---AAGEVLRKYFRQRVEIIDK--------EDHS---PVTIADRE 118
           |: :||  :     |   |: :|:   ||    : ||        : |:   ||| |||
Db       3 LSADRLAALDAFILDLNRASADVILPLFRADHGLEDKGAGKNLPRDTHAAFDPVTEADRG 62

Qy     119 AEEAMVSVILKSFPTHAIFGEENGWRCAEN--SADFVWVLDPIDGTKSFITGKPLFGTLI 176
           || |: ::| : || | : ||| |    |:   |:||||||||||::|| | ||: |||
Db      63 AEAAIRALIAQRFPDHGVIGEEYG----EDRPDAEFVWVLDPIDGTRAFIAGLPLWTTLI 118

Qy     177 ALLHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQE---ISVRSCNLLAQAYLYTTSPH- 232
           | | |:||:| | || : | :||  |    ::| |   | || |  :  | : || |
Db     119 GLRHEGRPVLGSIGQPYVNEIFIGHAGGARLVSGGEARPIRVRECANINDAVIATTDPDA 178

Qy     233 LFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGA 292
           |:     |:::||   |:   ||| ||||::| | :|:|:|:|:|:|| :|  : :|:||||
Db     179 CFDGAERGAWLQVRAAAKLARLGCDAYAYAMVAMGKMDMVIEAGLKSWDIEAAIPLIEGA 238

Qy     293 GGSITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           || :|:|||:    ||     |    :| :|| |   ||| :||
Db     239 GGMVTNWRGE----PV----GPNGGQMVISGDRRPLDEALVSLR 274


RESULT   6
AF3515
extragenic suppressor protein suhb [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C;Accession: AF3515
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AF3515
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-272 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL53289.1; PID:g17984172; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEII0048
A;Map position: II
C;Superfamily: suppressor protein suhB

  Query Match             26.8%;  Score 463.5;  DB 2;  Length 272;
  Best Local Similarity   41.3%;  Pred. No. 5.3e-29;
  Matches  116;  Conservative   47;  Mismatches   93;  Indels   25;  Gaps    9;

Qy      66 RAAAVGELATERLVEVA---QRAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAE 120
           ||   |   |  |:: |   : || || : | : |||  |: :|      ||| ||| ||
Db       4 RAPCFGNRRTGLLIDKAFFSEVAAAAAAQTLPR-FRQLTEVDNKYSVGFDPVTEADRAAE 62

Qy     121 EAMVSVILKSFPTHAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALL 179
           |: :|| ::|| | | ||| |   |||:    ||::||:|||::||:| |::|||: |
Db      63 RAIRAVIGRTFPYHGILGEEYG---AENTDRSHVWIIDPVDGTRAFISGLPVWGTLVGLT 119

Qy     180 HNGKPVIGVIDQPILRERWIGVDGKQTTLN-----GQEISVRSCNLLAQAYLYTTSPHLF 234
           :|    |:: ||   | :   ||    |       : : ||   :|  | |:||:| ||
Db     120 VDGDARAGMMSQPFTGELFYS-DGDGAYLQRGDGAPRRLCVRKNAVLEDATLFTTTPALF 178

Qy     235 EADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGG 294
           : |   || |: : |::  || ||||:|:|| ||||||||:||: ||  :|:|:|| |||
Db     179 KGDDRKAFDRLESAVRLSRYGVDCYAFAMLAGGFVDIVVEAGLQTYDIAALIPIIEQAGG 238

Qy     295 SITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           :|   |         ||      ::|||    :|: ||| |
Db     239 VVTRRDGG------PAEQ---GGDIVAAATPALHQAALDLL 270


RESULT   7
S74899
extragenic suppressor - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein sll1959
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S74899
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Ya
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocy
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74899
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <KAN>
A;Cross-references: EMBL:D90909; GB:AB001339; NID:g1652844; PIDN:BAA17860.1; PID:g16
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: suppressor protein suhB

  Query Match             22.9%;  Score 396.5;  DB 2;  Length 267;
  Best Local Similarity   36.9%;  Pred. No. 9.6e-24;
  Matches  101;  Conservative   49;  Mismatches   99;  Indels   25;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR----QRVEIIDKEDHSPVTIADREAEEAMVSVILKSF 131
           |: : :||: |  :||:| :|||    |    ||:   : || || |||:||| :|   |
Db       6 EQRLFIAQQLAAVSGEILIQYFRRSHLQGGTKIDQVS-AIVTQADEEAEQAMVDLIQAQF 64

Qy     132 PTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH-NGKPVIGVID 190
           |   :  || |   |  |  : |||||||| ||: | |:| ||| |:  : :||:|:
Db      65 PQDGVIREE-GKNIAGKSG-YTWVLDPIDGTSSFVRGLPIFATLIGLVDADMRPVLGIAH 122

Qy     191 QPILRERWIGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADA-EDAFIRV 245
           |||  :|| || |:|: :||            || | | : :|:| :|     :     :
Db     123 QPISGDRWQGVQGEQSNVNGIPLVNPYKASEINLTA-ACIVSTTPLMFTTPVQQQKMADI 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVD---IVVESGLKPYDFLSLVPVIEGAGGSITDWRGD 302
           : :   :| ||: |   |||:     ::||: |  ||| :|:|:: ||    |||:|
Db     182 YRQCQRTAFGGDCFNYLSAASGWTAMPLVIVEADLNFYDFCALIPILTGANYCFTDWQGK 241

Qy     303 KLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           :|         | |  |||: : ::| | |  |:
Db     242 EL--------TPESTEVVASPNPKLHSEILAFLQ 267


RESULT   8
AC0353
```

```
probable inositol monophosphatase family protein YPO2899 [imported] - Yersinia pestis (s
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AC0353
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0353
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92150.1; PID:g15980865; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2899
C;Superfamily: suppressor protein suhB


  Query Match             16.9%;  Score 293;  DB 2;  Length 267;
  Best Local Similarity   31.6%;  Pred. No. 1.3e-15;
  Matches   86;  Conservative   43;  Mismatches  111;  Indels   32;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || :: | :     ||   |  :  ||  ||:||  :: || ||:| |
Db       4 MLTIAIRAARKAGNLIAKNYETPDAVEASQKGSNDFVTNVDRDAEHLIIDVIRKSYPKHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |    |:  |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db      64 IISEECGELVGEDD-DVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCN-------LLAQAYLYTTSPHLFEADAEDAFIRVRNK 248
           | :    |:   |||  :  |  |        :||  : :    |          |:|||  |
Db     123 ELFTASRGQGAQLNGYRL--RGTNAKDLDGTILATGFPFKVKQH------APAYIRVVGK 174

Qy     249 V-----KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           :          |      | :|:| ||    | ||||:||      ::  :|| :||: |
Db     175 LFEQCADFRRTGSAALDLAYVAAGRVDGFFEIGLKPWDFAGGELLVRESGGIVTDFAGGH 234

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           |:         :| |:| ||: |: |   : |:
Db     235 NHF--------SSGNIV-AGNPRIVKSIVQAM 257


RESULT   9
C83169
extragenic suppressor protein SuhB PA3818 [imported] - Pseudomonas aeruginosa (strain PA
  Species: Pseudomonas aeruginosa
  Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
  Accession: C83169
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa [ 01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: C83169
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-271 <STO>
A;Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIDN:AAG07205.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3818
C;Superfamily: suppressor protein suhB


  Query Match             16.9%;  Score 292;  DB 2;  Length 271;
  Best Local Similarity   29.7%;  Pred. No. 1.6e-15;
  Matches   82;  Conservative   54;  Mismatches  106;  Indels   34;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEII---DKEDHSPVTIADREAEEAMVSVILKSFP 132
```

```
           :: :| |||  :|||::   ||  :|:::|   :|:    ||  || ||: :|: : |::|
Db       4 MLNIALRAARSAGELI---FRSIERLDVISVNEKDAKDYVTEVDRAAEQTIVAALRKAYP 60

Qy     133 THAIFGEENGW-RCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
           |||| ||| |:   :   ||::||:||:||| :|| | | |   ||  : |:    |:
Db      61 THAIMGEEGGFIEGSGEGADYLWVIDPLDGTTNFIHGVPHFAVSIACKYKGRLEHAVVLD 120

Qy     192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHL-FEADAEDAFIR-----V 245
           |: :| :    |:   |||: : |     |  | | |  |    : |  | ::      |
Db     121 PVRQEEFTASRGRGAALNGRRLRVSGRKSLEGALLGTGFPFRDNQIDNLDNYLNMFRSLV 180

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG---- 301
           :  |      | :|:|  |   | ||  :|  :   ::: ||| ::|: |
Db     181 GQTAGIRRAGAASLDLAYVAAGRYDAFWEFGLSEWDMAAGALLVQEAGGLVSDFTGSHEF 240

Qy     302 -DKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
            :| |              : ||: :  |  |  ::
Db     241 LEKGH-------------IVAGNTKCFKALLTTIQ 262


RESULT  10
T45317
monophophatase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 18-Feb-2000
C;Accession: T45317
R;Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z22864
A;Accession: T45317
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-255 <PAR>
A;Cross-references: EMBL:Z98271; PIDN:CAB11010.1
A;Experimental source: cosmid B1779
C;Genetics:
A;Note: MLCB1779.29
C;Superfamily: suppressor ; otein suhB


  Query Match             16.4%;  Score 283;  DB 2;  Length 255;
  Best Local Similarity   31.4%;  Pred. No. 7.6e-15;
  Matches   85;  Conservative   40;  Mismatches  102;  Indels   44;  Gaps    9;

Qy      81 VAQRAADAAGEVLRKYF-RQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   || |  :    |   : :  | | :||| |||  |  : :|: :  |   | ||
Db       2 LALTLADRADALISAREGALNLRVL ?P' TPVTDADRAVEADVRAVLGRERPKDGILGE 61

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW- 198
           | |      :     |::|||||||:|: | |:: :|||| :| | |||:  | |: ||
Db      62 EYGGTITFSGQQ--WIVDPIDGTKNFVRGVPVWASLIALLEDGVPSIGVVSAPALQRRWW 119

Qy     199 --------IGVDG--------KQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAF 242
                   : |||        :   ||   :|  | :  ||  |                | |
Db     120 AARGQGAFVAVDGVPRRLAVSEVADLNSASLSFSSLSGWAQRGL-----------RDRF 167

Qy     243 IRVRNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
           : : : | :|  || |  :| ||| | :|:  |  :  :|  :|  |:  ||| :|   |
Db     168 LELTDAVWRVRAYG-DFLSYCLLAEGAIDVAAEPKVSVWDLAALDIVVREAGGVLTGLDG 226

Qy     302 DKLHWPVTAESRPTSFNVVAAGDARVHKEAL 332
                   |   : ||  : |:|:| |
Db     227 ---------TPGPHGGSAVAT-NGRLHQEVL 247


RESULT  11
C70470
myo-inositol-1(or 4)-monophosphatase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: C70470
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
```

Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: C70470
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-264 <AQF>
A;Cross-references: GB:AE000766; GB:AE000657; NID:g2984216; PIDN:AAC07753.1; PID:g298422
A;Experimental source: strain VF5
C;Genetics:
A;Gene: imp2
C;Superfamily: suppressor protein suhB

Query Match 16.1%; Score 279; DB 1; Length 264;
Best Local Similarity 28.5%; Pred. No. 1.6e-14;
Matches 75; Conservative 54; Mismatches 106; Indels 28; Gaps 7;

```
Qy      76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  |||| ||
Db       5 KKYLEVAKIAALAGGQVLKENFGKVKKENIEEKGEKDFVSYVDKTSEERIKEVILKFFPD 64

Qy     134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           | : ||| |   || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 HEVVGEEMG---AEGSGSEYRWFIDPLDGTKNYINGFI.FAVSVGLVKGEEPIVGAVYLP 121

Qy     193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRV 245
              : : |  |    :||: | |:    |  | :    |      |        :| |  |
Db     122 YFDKLYWGAKGLGAYVNGKRIKVKDNESLKHAGVVYGFPSRSRRDISIYLNIFKDVFYEV 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
              :: |  |       ::| |  | ::|  :||:|  : : ::: |||
Db     182 -GSMRRP--GAAAVDLCMVAEGIFDGMMEFEMKPWDITAGLVILKEAGG---------- 227

Qy     306 WPVTAESRPTSFNVVAAGDARVH 328
              |    |   : : ||:  :|
Db     228 -VYTLVGEPFGVSDIIAGNKALH 249
```

RESULT 12
B70646
probable monophosphatase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: B70646
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K ; Skelton, S.; Squares, S.
ature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70646
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-260 <COL>
A;Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CAB06277.1; PID:g1781230
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3137
C;Superfamily: suppressor protein suhB

Query Match 15.8%; Score 273.5; DB 2; Length 260;
Best Local Similarity 32.5%; Pred. No. 4.4e-14;
Matches 88; Conservative 42; Mismatches 104; Indels 37; Gaps 11;

```
Qy      81 VAQRAADAAGEVLR-KYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   || | |: | ::    : |  | | :||| |||  |  :   : :  |   : ||
Db       8 LALALADRADELTRVRFGALDLRIDTKPDLTPVTDADRAVESDVRQTLGRDRPGDGVLGE 67

Qy     140 ENGWRCAENSADFV---WVLDPIDGTKSFITGKPLFGT.IALLHNGKPVIGVIDQPILRE 196
           | |      |  |    |::||||||:|: | |:: :|||| :| | :||:  | |:
```

```
Db      68 EFG-----GSTTFTGR(WIVDPIDGTKNFVRGVPVWASLIALLEDGVPSVGVVSAPALQR 122

Qy     197 RWIGVDGKQ--TTLNG---QEISVRSCNLLAQAYLYTTS------PHLFEADAEDAFIRV 245
           ||    |:    :::|     :|| |   |  | | :|      | |      : || :
Db     123 RWWAARGRGAFASVDGARPHRLSVSSVAELHSASLSFSSLSGWARPGL-----RERFIGL 177

Qy     246 RNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKL 304
           : | :|  || |  :| |:| | |||  |  :  :|  :|  |:  ||| :|   |
Db     178 TDTVWRVRAYG-DFLSYCLVAEGAVDIAAEPQVSVWDLAALDIVVREAGGRLTSLDG--- 233

Qy     305 HWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
                 : |   : ||  :  :| | |  |
Db     234 ------VAGPHGGSAVAT-NGLLHDEVLTRL 257
```

RESULT 13
F64103
suppressor protein suhB - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C;Accession: F64103
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavag
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidma
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vent
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: F64103
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-267 <TIGR>
A;Cross-references: GB:U32775; GB:L42023; NID:g1573951; PIDN:AAC22595.1; PID:g157395
C;Genetics:
A;Gene: suhB
C;Superfamily: suppressor protein suhB

Query Match 15.7%; Score 272; DB 1; Length 267;
Best Local Similarity 30.8%; Pred. No. 5.9e-14;
Matches 84; Conservative 44; Mismatches 113; Indels 32; Gaps 8;

```
Qy      78 LVEVAQRAADAAGEVLRKYFRQR--VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || |: | : :|  :|   |  :  ||  |: :|  :: || ||:| |
Db       4 MLNIAIRAARKAGNVIAKNYERRDAIESTQKGINDYVTNVDKASEAEIIEVIRKSYPDHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |       :|  |::|||:|||::|:|| | |   ||:    :  :||:  ||
Db      64 IITEETG-AIEGKDSDVQWIIDPLDGTRNFMTGLPHFSVSIAVRVKNRTEVGVVYDPIRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPH----------LFEADAEDAF-I 243
           | :  | |:   ||   : | |   :  : | |  |              :  |  |||
Db     123 ELFTAVRGEGAKLNEVRLRVDSKREIQGSILATGFPFKQPKLMPAQFAMMNALIEDAADF 182

Qy     244 RVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           |       : |    ||    :||  :|   | ||| :|  :   ::  ||| : |:
Db     183 RRTGSAALDL----CY----VASNRIDGYFEMGLKAWDCAAGDLIVREAGGLVCDFD--- 231

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                 |   | |::||  :|| || |: :|
Db     232 -----AGNSYLRSGNIIAA-PSRVIKEMLNKIR 258
```

RESULT 14
T07795
myo-inositol-1(or 4)-monophosphatase (EC 3.1.3.25) isoform 3 - tomato
N;Alternate names: myo-inositol-1-phosphatase
C;Species: Lycopersicon esculentum (tomato)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C;Accession: T07795
R;Gillaspy, G.E.; Keddie, J.S.; Oda, K.; Gruissem, W.
Plant Cell 7, 2175-2185, 1995
A;Title: Plant inositol monophosphatase is a lithium-sensitive enzyme encoded by a m

A;Reference number: Z16138; MUID:96351935
A;Accession: T07795
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-268 <GIL>
A;Cross-references: EMBL:U39059; NID:g1098970; PIDN:AAB19031.1; PID:g1098971
A;Experimental source: strain VFNT Cherry
C;Genetics:
A;Gene: IMP3
C;Function:
A;Description: removes phosphate from myo-inositol-1-phosphate and myo- inositol 4-phosp
A;Pathway: inositol phosphate metabolism
C;Superfamily: suppressor protein suhB
C;Keywords: phosphoric monoester hydrolase

Query Match 15.3%; Score 265.5; DB 2; Length 268;
Best Local Similarity 30.2%; Pred. No. 1.9e-13;
Matches 83; Conservative 48; Mismatches 117; Indels 27; Gaps 8;

```
Qy      74 ATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           : |: ::||  ||  |||::|: | :   :  |     ||  |:  |: : : : : ||:
Db       6 SVEQFLDVAVEAAKKAGEIIREGFYKTKHVEHKGMVDLVTETDKACEDFIFNHLKQRFPS 65

Qy     134 HAIFGEENGWRCA--ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
           |   |||    |   | : :  |::||:|||  :|: | |     | |    || :||:
Db      66 HKFIGEETTAACGNFELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIEKKPTVGVVYN 125

Qy     192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT---TSPHLFEADA-----EDAFI 243
           ||: | : |:|||   |||: | | | : | :| | |   |:      ||
Db     126 PIIDELFTGIDGKGAFLNGKPIKVSSQSELVKALLATEAGTNRDKLVVDATTGRINSLLF 185

Qy     244 RVRNKVKVPLYGCDCYAYAL--LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
           :||:     |  |   |  :| | :|:  |     |:|      ::: ||| : |
Db     186 KVRS-----LRMCGSCALNLCGVACGRLDLFYELEFGGPWDVAGGAVIVKEAGGFVFDPS 240

Qy     301 GDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           | :  : :||         ||| :| :    : ||
Db     241 GSE--FDLTARR-------VAATNAHLKDAFIKAL 266
```

RESULT 15
D65030
myo-inositol-1(or 4)-monophosphatase (EC 3.1.3.25) [validated] - Escherichia coli
N;Alternate names: suppressor protein suhB
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: D65030; A35158
Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617
A;Accession: D65030
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-267 <BLAT>
A;Cross-references: GB:AE000339; GB:U00096; NID:g1788870; PIDN:AAC75586.1; PID:g1788882;
A;Experimental source: strain K-12, substrain MG1655
R;Yano, R.; Nagai, H.; Shiba, K.; Yura, T.
J. Bacteriol. 172, 2124-2130, 1990
A;Title: A mutation that enhances synthesis of sigma(32) and suppresses temperature-sens
A;Reference number: A35158; MUID:90202735
A;Accession: A35158
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-140,'L',142-267 <YAN>
A;Cross-references: GB:M34828; NID:g147898; PIDN:AAA67506.1; PID:g147899
C;Genetics:
A;Gene: suhB
C;Complex: monomer [validated, MUID:20213266]
C;Function:
A;Description: EC 3.1.3.25 [validated, MUID:20213266]
A;Note: inositol-1-phosphatase activity is neither sufficient nor required for comple
C;Superfamily: suppressor protein suhB
C;Keywords: phosphoric monoester hydrolase

Query Match 14.9%; Score 258; DB 1; Length 267;
Best Local Similarity 29.6%; Pred. No. 7.5e-13;
Matches 79; Conservative 44; Mismatches 124; Indels 20; Gaps 7;

```
Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || :: | :     ||   |  :  ||  |: ||  ::  | ||:| |
Db       4 MLNIAVRAARKAGNLIAKNYETPDAVEASQKGSNDFVTNVDKAAEAVIIDTIRKSYPQHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  ||:|         |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db      64 IITEESG-ELEGTDQDVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKV----- 249
           | :    |:   ||| :   :   |   | |  |  |:|      :| :  |:
Db     123 ELFTATRGQGAQLNGYRLRGSTARDLDGTILATGFP--FKAKQYATTYINIVGKLFNECA 180

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
                 |      | :|:| ||   | ||:|:|| :   ::  |||| ::|: |   :: :|
Db     181 DFRRTGSAALDLAYVAAGRVDGFFEIGLRPWDFAAGELLVREAGGIVSDFTGGH-NYMLT 239

Qy     310 AESRPTSFNVVAAGDARVHKEALDALR 336
                   : ||: || | ·|  :|
Db     240 GN--------IVAGNPRVVKAMLANMR 258
```

Search completed: May 30, 2002, 17:15:56
Job time: 207 sec

```
;  GENERAL INFORMATION:
;       APPLICANT: Detera-Wadleigh, Sevilla D.
;                  Gershon, Elliot S.
;                  Badner, Judith A.
;                  Goldin, Lynn R.
;                  Berrettini, Wade H.
;                  Yoshikawa, Takeo
;                  Sanders, Alan R.
;                  Esterling, Lisa E.
;       TITLE OF INVENTION: Chromosomal Markers and Diagnostic
;                           Tests for Manic-Depressive Illness
;       NUMBER OF SEQUENCES: 197
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Townsend and Townsend and Crew LLP
;            STREET. Two Embarcadero Center, Eighth Floor
;            CITY: San Francisco
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94111-3834
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/091,952A
;            FILING DATE: 19-Apr-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 60/029,278
;            FILING DATE: 28-OCT-1996
;            APPLICATION NUMBER: PCT/US97/19381
;            FILING DATE: 28-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Smith, Timothy L.
;            REGISTRATION NUMBER: 35,367
;            REFERENCE/DOCKET NUMBER: 015280-297100US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 576-0200
;            TELEFAX: (415) 576-0300
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 25:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 285 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       FEATURE:
;            NAME/KEY:  mat_peptide
;            LOCATION:  1...285
;            OTHER INFORMATION:  Xenopus IMP
;       SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-09-091-952A-25

  Query Match            12.1%;  Score 209;  DB 4;  Length 285;
  Best Local Similarity  25.8%;  Pred. No. 1.9e-14;
  Matches   62;  Conservative   48;  Mismatches  104;  Indels   26;  Gaps    6;

Qy      81 VAQRAADAAGEVLRKYFRQRVEIIDKEDHSP---VTIADREAEEAMVSVILKSFPTHAIF 137
           :|   |  || |:    :: | |: |   :|   ||  |:: || ::| | : :|:|:
Db      12 LAVSIARKAGSVVCAALKEDVSIMVKTSLAPADLVTA?DQKVEEMIISSIKEKYPSH?FI 71

Qy     138 GEEN-----GWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           |||:     |    :|     |::|||||| :|:   |      |    : :   ||:
Db      72 GEESVAAGAGSTLTDNP---TWIIDPIDGTTNFVHRFPFVAVSIGFAVHKQVEFGVVYSC 128

Qy     193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIR------VR 246
           :  : : |  || :  |||:: |     : :: : |       ::    ||:      :
Db     129 VEDKMYTGRKGKGSFCNGQKLQVSGQKDITKSMIITE----LGSNRNPEFIKTVSLSNME 184

Qy     247 NKVKVPLYGCDCYAYA-----LLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
             : :|::|      |     |:|:|  |   | ||  :|  :   ::  |||:| |  |
```

```
Db     185 RLLCIPIHGIRAVGTAAVNMCLVATGGADAYYEMGLHCWDMAAASVIVTEAGGTILDATG 244


RESULT 5
US-09-134-001C-3445
; Sequence 3445, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3445
;   LENGTH: 281
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3445

  Query Match            12.0%;  Score 208;  DB 4;  Length 281;
  Best Local Similarity  24.5%;  Pred. No. 2.4e-14;
  Matches   64;  Conservative   51;  Mismatches  136;  Indels   10;  Gaps    4;

Qy      75 TERLVEVAQRAADAAGEVLRKYFRQRVEIIDKED-HSPVTIADREAEEAMVSVILKSFPT 133
           |  | : |:     ||  :|   :: ::|  | : :  ||  |:  |  :   || ::|
Db       8 TMALYDFAKGLILEAGNKVRLMMQEELDIKTKSNPNDLVTNVDKATENYLYETILHNYPD 67

Qy     134 HAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPI 193
           | : |||      |     :||:|||||| :|:  |  |   | : |:|||  | :   :
Db      68 HQVIGEEGHGHNLEYLKGVIWVIDPIDGTLNFVHQKENFAISIGIYHDGKPYAGFVYDVM 127

Qy     194 LRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVK-VP 252
               :    |:    |  :: :     | :: :       | :    | |  : |: :
Db     128 KDVLYHAKVGQGAFENTHKLEMIQNTELKRSIIGINPNWLSKPILSDIFSSIVNEARSAR 187

Qy     253 LYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAES 312
             ||        :| | :   :   |:|:||   : ::   ||  |:  ||||  :     :
Db     188 AYGSAALEIISVAKGQLAAYLTPRLQPWDFAGGLLILNEVGGIGTNLLGDKLDF-----N 242

Qy     313 RPTSFNVVAAGDARVHKEALD 333
           :|   | :   :  :|:| |:
Db     243 QP---NSILIANPSLHREILN 260


RESULT 6
US-08-347-471-4
; Sequence 4, Application US/08347471
; Patent No. 5710015
;  GENERAL INFORMATION:
;    APPLICANT:  McAllister, George
;    APPLICANT:  Whiting, Paul J.
;    TITLE OF INVENTION:  cDNA CLONING OF INOSITOL MONOPHOSPHATASE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Robert J. No. 5710015th
;      STREET:  P.O. Box 2000, 126 E. Lincoln Avenue
;      CITY:  Rahway
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/347,471
```

```
;       FILING DATE:  08-DECEMBER-94
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 5710015th, Robert J.
;       REGISTRATION NUMBER:  27,366
;       REFERENCE/DOCKET NUMBER:  T-1160
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908) 594-7262
;       TELEX:  (908) 594-4720
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  270 amino acids
;       TYPE:  amino acid
;       TOPOLOCY:  linear
;     MOLECULE TYPE:  protein
US-08-347-471-4

  Query Match             11.5%;  Score 199.5;  DB 1;  Length 270;
  Best Local Similarity   25.2%;  Pred. No. 1.9e-13;
  Matches   67;  Conservative   51;  Mismatches  111;  Indels   37;  Gaps   10;

Qy      86 ADAAGEVLRKYFRQRVEIIDKEDHSP---VTIADREAEEAMVSVILKSFPTHAIFGEENG 142
           |  |||::|   : ::::: |   ||   ||: |:: |: ::| | : :| |:  |||:
Db       9 ARQAGEMIRVALKNKMDVMIKS--SPADLVTVTDQKV KMLMSSIKEKYPYHSFIGE.:- 65

Qy     143 WRCAENSADF----VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
              |     |     |::|||||| :|:   |     |  : | :   ||:   :  : :
Db      66 -VAAGEKTVFTEQPTWIIDPIDGTTNFVHRFPFVAVSIGFVVNKEMEFGVVYSCVEDKMY 124

Qy     199 IGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRV-----RNKVKVPL 253
           |  ||    |||:: |     : :: | |       :  :   :|:          :|:
Db     125 TGRKGKGAFCNGQKLRVSQQEDITKSLLVTE----LGSSRKPETLRIVLSNMERLCSIPI 180

Qy     254 YGCDCYAYA-----LLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPV 308
           :|      |     |:|:|  |   | |:  :|      ::  ||| : |  |     |
Db     181 HGIRSVGTAAVNMCLVATGGADAYYEMGIHCWDMAGAGIIVIEAGGVLLDVTGG----PF 236

Qy     309 TAESRPTSFNVVAAGD----ARVHKE 330
              ||     ::|| :     |: ||
Db     237 DLMSR----RIIAASNIALAERIAKE 258


RESULT 7
US-09-091-952A-17
; Sequence 17, Application US/09091952A
; Patent No. 6458532
;   GENERAL INFORMATION:
;       APPLICANT: Detera-Wadleigh, Sevilla D.
;                  Gershon, Elliot S.
;                  Badner, Judith A.
;                  Goldin, Lynn R.
;                  Berrettini, Wade H.
;                  Yoshikawa, Takeo
;                  Sanders, Alan R.
;                  Esterling, Lisa E.
;       TITLE OF INVENTION: Chromosomal Markers and Diagnostic
;                           Tests for Manic-Depressive Illness
;       NUMBER OF SEQUENCES: 197
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Townsend and Townsend and Crew LLP
;            STREET: Two Embarcadero Center, Eighth Floor
;            CITY: San Francisco
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94111-3834
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/091,952A
;            FILING DATE: 19-Apr-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 60/029,278
;            FILING DATE: 28-OCT-1996
;            APPLICATION NUMBER: PCT/US97/19381
;            FILING DATE: 28-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Smith, Timothy L.
;            REGISTRATION NUMBER: 35,367
;            REFERENCE/DOCKET NUMBER: 015280-297100US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 576-0200
;            TELEFAX: (415) 576-0300
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 17:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 285 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       FEATURE:
;            NAME/KEY:  mat_peptide
;            LOCATION:  1...288
;            OTHER INFORMATION:  IMP.18p myo-inositol monophosphatase
;       SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-091-952A-17

  Query Match             11.4%;  Score 196.5;  DB 4;  Length 288;
  Best Local Similarity   26.4%;  Pred. No. 4.4e-13;
  Matches   67;  Conservative   40;  Mismatches  132;  Indels   15;  Gaps    5;

Qy      63 AGRRAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSP-VTIADREAEE 121
           :|   ||:     |   : | : |  ||:::||   :   :  |   :  ||  |   |:
Db       4 SGEDQAALAAGPWEECFQAAVQLALRAGQIIRKALTEEKRVSTKTSAADLVTETDHLVED 63

Qy     122 AMVSVILKSFPTHAIFGEE---NGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIAL 178
           ::| : : ||:|    ||   :| :|    :   |::||||| :|:   |     |
Db      64 LIISELRERFPSHRFIAEEAAASGAKCVLTHSP-TWIIDPIDGTCNFVHRFPTVAVSIGF 122

Qy     179 LHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT--------TS 230
             :   |||         : |  |:    ||| : |     |::| : |          :
Db     123 AVRQELEFGVIYHCTEERLYTGRRGRGAFCNGQRLRVSGETDLSKALVLTEIGPKRDPAT 182

Qy     231 PHLFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIE 290
             || :: |       : |:|   |     |   ||||  |   : ||  :|  :   :|
Db     183 LKLFLSNMERLLHAKAHGVRV--IGSSTLALCHLASGAADAYYQFGLHCWDLAAATVIIR 240

Qy     291 GAGGSITDWRGDKL 304
           |||  : |  |  |
Db     241 EAGGIVIDTSGGPL 254


RESULT 8
US-08-347-471-2
; Sequence 2, Application US/08347471
; Patent No. 5710015
;  GENERAL INFORMATION:
;    APPLICANT:  McAllister, George
;    APPLICANT:  Whiting, Paul J.
;    TITLE OF INVENTION:  cDNA CLONING OF INOSITOL MONOPHOSPHATASE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Robert J. No. 5710015th
;      STREET:  P.O. Box 2000, 126 E. Lincoln Avenue
;      CITY:  Rahway
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/347,471
;       FILING DATE:  08-DECEMBER-94
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  No. 5710015th, Robert J.
;       REGISTRATION NUMBER:  27,366
;       REFERENCE/DOCKET NUMBER:  T-1160
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908) 594-7262
;       TELEX:  (908) 594-4720
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  277 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-347-471-2

  Query Match             11.2%;  Score 193.5;  DB 1;  Length 277;
  Best Local Similarity   25.5%;  Pred. No. 8.9e-13;
  Matches   70;  Conservative   51;  Mismatches  120;  Indels   33;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSP---VTIADREAEEAMVSVILKSFP 132
           :  :: |   |  ||||: :  :  : :: |   ||   ||  |:: |: ::| | : :|
Db       6 QECMDYAVTLARQAGEVVCEAIKNEMNVMLKS--SPVDLVTATDQKVEKMLISSIKEKYP 63

Qy     133 THAIFGEENGWRCAENSA---DFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           :|:  |||:     | |    :  |::|||||| :|:   |     |    | |   ||:
Db      64 SHSFIGEES-VAAGEKSILTDNPTWIIDPIDGTTNFVHRFPFVAVSIGFAVNKKIEFGVV 122

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           :  : :    ||    |||:: |     : :: | |    |  :   :    | : :
Db     123 YSCVEGKMYTARKGKGAFCNGQKLQVSQQEDITKSLLVT---ELGSSRTPETVRMVLSNM 179

Qy     250 K----VPLYGCDCYAYA-----LLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300
           :    :|::|     |     |:|:|  |   | |:  :|      :. |||  : |
Db     180 EKLFCIPVHGIRSVGTAAVNMCLVATGGADAYYEMGIHCWDVAGAGIIVTEAGGVLMDVT 239

Qy     301 GDKLHWPVTAESRPTSFNVVAAGD----ARVHKE 330
           |     |    ||     |:|| :     |: ||
Db     240 GG----PFDLMSR----RVIAANNRILAERIAKE 265


RESULT 9
US-09-091-952A-26
; Sequence 26, Application US/09091952A
  Patent No. 6458532
;   GENERAL INFORMATION:
;       APPLICANT: Detera-Wadleigh, Sevilla D.
;                  Gershon, Elliot S.
;                  Badner, Judith A.
;                  Goldin, Lynn R.
;                  Berrettini, Wade H.
;                  Yoshikawa, Takeo
;                  Sanders, Alan R.
;                  Esterling, Lisa E.
;       TITLE OF INVENTION: Chromosomal Markers and Diagnostic
;                           Tests for Manic-Depressive Illness
;       NUMBER OF SEQUENCES: 197
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Townsend and Townsend and Crew LLP
;            STREET: Two Embarcadero Center, Eighth Floor
;            CITY: San Francisco
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94111-3834
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/091,952A
;            FILING DATE: 19-Apr-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 60/029,278
;            FILING DATE: 28-OCT-1996
;            APPLICATION NUMBER: PCT/US97/19381
;            FILING DATE: 28-OCT-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Smith, Timothy L.
;            REGISTRATION NUMBER: 35,367
;            REFERENCE/DOCKET NUMBER: 015280-297100US
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 576-0200
;            TELEFAX: (415) 576-0300
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 26:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 277 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: <Unknown>
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       FEATURE:
;            NAME/KEY:  mat_peptide
;            LOCATION:  1...277
;            OTHER INFORMATION:  rat IMP
;       SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-091-952A-26

  Query Match             11.2%;  Score 193.5;  DB 4;  Length 277;
  Best Local Similarity   25.5%;  Pred. No. 8.9e-13;
  Matches   70;  Conservative   51;  Mismatches  120;  Indels   33;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSP---VTIADREAEEAMVSVILKSFP 132
           :  :: |   |  ||||: :  :  : :: |   ||   ||  |:: |: ::| | : :|
Db       6 QECMDYAVTLARQAGEVVCEAIKNEMNVMLKS--SPVDLVTATDQKVEKMLISSIKEKYP 63

Qy     133 THAIFGEENGWRCAENSA---DFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           :|:  |||:     | |    :  |::|||||| :|:   |     |    | |   ||:
Db      64 SHSFIGEES-VAAGEKSILTDNPTWIIDPIDGTTNFVHRFPFVAVSIGFAVNKKIEFGVV 122

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           :  : :    ||    |||:: |     : :: | |    |  :   :    | : :
Db     123 YSCVEGKMYTARKGKGAFCNGQKLQVSQQEDITKSLLVT---ELGSSRTPETVRMVLSNM 179

Qy     250 K----VPLYGCDCYAYA-----LLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300
           :    :|::|     |     |:|:|  |   | |:  :|      ::  ||| : |
Db     180 EKLFCIPVHGIRSVGTAAVNMCLVATGGADAYYEMGIHCWDVAGAGIIVTEAGGVLMDVT 239

Qy     301 GDKLHWPVTAESRPTSFNVVAAGD----ARVHKE 330
           |     |    ||     |:|| :     |: ||
Db     240 GG----PFDLMSR----RVIAANNRILAERIAKE 265


RESULT 10
US-09-091-952A-28
; Sequence 28, Application US/09091952A
; Patent No. 6458532
;   GENERAL INFORMATION:
;       APPLICANT: Detera-Wadleigh, Sevilla D.
;                  Gershon, Elliot S.
;                  Badner, Judith A.
;                  Goldin, Lynn R.
;                  Berrettini, Wade H.
;                  Yoshikawa, Takeo
;                  Sanders, Alan R.
;                  Esterling, Lisa E.
;       TITLE OF INVENTION: Chromosomal Markers and Diagnostic
;                           Tests for Manic-Depressive Illness
;       NUMBER OF SEQUENCES: 197
;       CORRESPONDENCE ADDRESS:
```

```
;           ADDRESSEE: Townsend and Townsend and Crew LLP
;           STREET: Two Embarcadero Center, Eighth Floor
;           CITY: San Francisco
;           STATE: CA
;           COUNTRY: USA
;           ZIP: 94111-3834
;         COMPUTER READABLE FORM:
;           MEDIUM TYPE: Diskette
;           COMPUTER: IBM Compatible
;           OPERATING SYSTEM: DOS
;           SOFTWARE: FastSEQ for Windows Version 2.0
;         CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/091,952A
;           FILING DATE: 19-Apr-1999
;           CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 60/029,278
;           FILING DATE: 28-OCT-1996
;           APPLICATION NUMBER: PCT/US97/19381
:           FILING DATE: 28-OCT-1997
;         ATTORNEY/AGENT INFORMATION:
;           NAME: Smith, Timothy L.
;           REGISTRATION NUMBER: 35,367
;           REFERENCE/DOCKET NUMBER: 015280-297100US
;         TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (415) 576-0200
;           TELEFAX: (415) 576-0300
;           TELEX: <Unknown>
;     INFORMATION FOR SEQ ID NO: 28:
;         SEQUENCE CHARACTERISTICS:
;           LENGTH: 277 amino acids
;           TYPE: amino acid
;           STRANDEDNESS: <Unknown>
;           TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         FEATURE:
;           NAME/KEY: mat_peptide
;           LOCATION: 1...277
;           OTHER INFORMATION: human IMP
;         SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-091-952A-28

  Query Match             11.2%;  Score 193.5;  DB 4;  Length 277;
  Best Local Similarity   25.5%;  Pred. No. 8.9e-13;
  Matches   70;  Conservative   51;  Mismatches  120;  Indels   33;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSP---VTIADREAEEAMVSVILKSFP 132
           :  :: |   |  ||||: :  :  : :: |   ||   ||  |:: |: ::| | : :|
Db       6 QECMDYAVTLARQAGEVVCEAIKNEMNVMLKS--SPVDLVTATDQKVEKMLISSIKEKYP 63

Qy     133 THAIFGEENGWRCAENSA---DFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           :|:  |||:     | |    :  |::|||||| :|:   |     |    | |   ||:
Db      64 SHSFIGEES-VAAGEKSILTDNPTWIIDPIDGTTNFVHRFPFVAVSIGFAVNKKIEFGVV 122

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           :  : :    ||    |||:: |     : :: | |    |  :   :    | : :
Db     123 YSCVEGKMYTARKGKGAFCNGQKLQVSQQEDITKSLLVT---ELGSSRTPETVRMVLSNM 179

Qy     250 K----VPLYGCDCYAYA-----LLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300
           :    :|::|      |     |:|:|  |   | |:  :|      ::  ||| : |
Db     180 EKLFCIPVHGIRSVGTAAVNMCLVATGGADAYYEMGIHCWDVAGAGIIVTEAGGVLMDVT 239

Qy     301 GDKLHWPVTAESRPTSFNVVAAGD----ARVHKE 330
           |     |    ||     |:|| :     |: ||
Db     240 GG----PFDLMSR----RVIAANNRILAERIAKE 265


RESULT 11
US-08-461-731-2
; Sequence 2, Application US/08461731
; Patent No. 5716806
;  GENERAL INFORMATION:
;    APPLICANT:  MEISSNER, ET AL.
;    TITLE OF INVENTION:  Human Inositol Monophosphatase H1
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/461,731
;      FILING DATE:  June 5, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/10465
;      FILING DATE:  16 SEP 94
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-316
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  265 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-08-461-731-2

  Query Match             10.0%;  Score 173;  DB 1;  Length 265;
  Best Local Similarity   24.8%;  Pred. No. 1.4e-10;
  Matches   67;  Conservative   35;  Mismatches  126;  Indels   42;  Gaps    7;

Qy      91 EVLRKYFRQRVEIIDKEDHSP-VTIADREAEEAMVSVILKSFPTHAIFGEE---NGWRCA 146
           |::||   :   :  |   :  ||  |  |: ::| : : ||:|    ||   :| :|
Db       9 EIIRKALTEEKRVSTKTSAADLVTETDHLVEDLIISELRERFPSHRFIAEEAAASGAKCV 68

Qy     147 ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQT 206
           :   |::|||||| :|:   |     |       :  |||        : |  |:
Db      69 LTHSP-TWIIDPIDGTCNFVHRFPIVAVSIGFAVRQELEFGVIYHCTEERLYTGRRGRGA 127

Qy     207 TLNGQEISVRSCNLLAQAYLYT---------------TSPHLFEADAEDAFIRVRNKVK 250
           |||  : |     |::| : |                    |  | |
Db     128 FCNGQRLRVSGETDLSKALVLTEIGPKRDPATLKLFLSNMGRLLHAKAHG---------- 177

Qy     251 VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
           | : |    |   ||||  |   : ||  :|  :   :|  ||| : |  |  |   |
Db     178 VRVIGSSTLALCHLASGAADAYYQFGLHCWDLAAATVIIREAGGIVIDTSGGPLDLMVC- 236

Qy     311 ESRPTSFNVVAAGD---ARVHKEALDALRW 337
                 ||||     | :  :||  : :
Db     237 -------RVVAASTREMAMLIAQALQTINY 259


RESULT 12
US-09-002-072B-2
; Sequence 2, Application US/09002072B
; Patent No. 5955339
;  GENERAL INFORMATION:
;    APPLICANT:  Meissner, Paul et al.
;    TITLE OF INVENTION:  Human Inositol Monophosphatase H1
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciencs, Inc.
```

```
;       STREET:  9410 Key West Avenue
;       CITY:  Rockville
;       STATE:  MD
;       COUNTRY:  USA
;       ZIP:  20850
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/002,072B
;       FILING DATE:  09-FEB-1998
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/461,731
;       FILING DATE:  05-JUN-1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US94/10465
;       FILING DATE:  16-SEP-1994
      ATTORNEY/AGENT INFORMATION:
;       NAME:  Brookes, A. Anders
;       REGISTRATION NUMBER:  36,373
;       REFERENCE/DOCKET NUMBER:  PF134D1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  301-309-8504
;       TELEFAX:  301-309-8439
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  265 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-002-072B-2

  Query Match           10.0%;  Score 173;  DB 2;  Length 265;
  Best Local Similarity  24.8%;  Pred. No. 1.4e-10;
  Matches   67;  Conservative   35;  Mismatches  126;  Indels   42;  Gaps    7;

Qy      91 EVLRKYFRQRVEIIDKEDHSP-VTIADREAEEAMVSVILKSFPTHAIFGEE---NGWRCA 146
           |::||   :   :  |   :  ||  |   |: ::| : : ||:|    ||   :| :|
Db       9 EIIRKALTEEKRVSTKTSAADLVTETDHLVEDLIISELRERFPSHRFIAEEAAASGAKCV 68

Qy     147 ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQT 206
             :   |::|||||| :|:  |     |      :   |||         : |  |:
Db      69 LTHSP-TWIIDPIDGTCNFVHRFPTVAVSIGFAVRQELEFGVIYHCTEERLYTGRRGRGA 127

Qy     207 TLNGQEISVRSCNLLAQAYLYT----------------TSPHLFEADAEDAFIRVRNKVK 250
             ||| : |     |::| : |                      |  | |
Db     128 FCNGQRLRVSGETDLSKALVLTEIGPKRDPATLKLFLSNMGRLLHAKAHG---------- 177

Qy     251 VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
           | : |    |   ||||  |   : ||  :|  :   :|  ||| : |  |  |   |
Db     178 VRVIGSSTLALCHLASGAADAYYQFGLHCWDLAAATVIIREAGGIVIDTSGGPLDLMVC- 236

Qy     311 ESRPTSFNVVAAGD---ARVHKEALDALRW 337
                 ||||     | :  :||  : :
Db     237 -------RVVAASTREMAMLIAQALQTINY 259


RESULT 13
US-09-314-199-2
; Sequence 2, Application US/09314199
; Patent No. 6130051
;  GENERAL INFORMATION:
;    APPLICANT:  MEISSNER, ET AL.
;    TITLE OF INVENTION:  Human Inositol Monophosphatase H1
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5 1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/314,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/002,072
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-316
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  265 AMINO ACIDS
;      TYPE:  AMINO ACID
;      STRANDEDNESS:
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  PROTEIN
US-09-314-199-2

  Query Match           10.0%;  Score 173;  DB 4;  Length 265;
  Best Local Similarity  24.8%;  Pred. No. 1.4e-10;
  Matches   67;  Conservative   35;  Mismatches  126;  Indels   42;  Gaps    7;

Qy      91 EVLRKYFRQRVEIIDKEDHSP-VTIADREAEEAMVSVILKSFPTHAIFGEE---NGWRCA 146
           |::||   :   :  |   :  ||  |   |: ::| : : ||:|    ||   :| :|
Db       9 EIIRKALTEEKRVSTKTSAADLVTETDHLVEDLIISELRERFPSHRFIAEEAAASGAKCV 68

Qy     147 ENSADFVWVLDPIDG KSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQT 206
             :   |::|||||| :|:  |     |      :   |||         : |  |:
Db      69 LTHSP-TWIIDPIDGTCNFVHRFPTVAVSIGFAVRQELEFGVIYHCTEERLYTGRRGRGA 127

Qy     207 TLNGQEISVRSCNLLAQAYLYT----------------TSPHLFEADAEDAFIRVRNKVK 250
             ||| : |     |::| : |                      |  | |
Db     128 FCNGQRLRVSGETDLSKALVLTEIGPKRDPATLKLFLSNMGRLLHAKAHG---------- 177

Qy     251 VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
           | : |    |   ||||  |   : ||  :|  :   :|  ||| : |  |  |   |
Db     178 VRVIGSSTLALCHLASGAADAYYQFGLHCWDLAAATVIIREAGGIVIDTSGGPLDLMVC- 236

Qy     311 ESRPTSFNVVAAGD---ARVHKEALDALRW 337
                 ||||     | :  :||  : :
Db     237 -------RVVAASTREMAMLIAQALQTINY 259


RESULT 14
US-09-314-198-2
; Sequence 2, Application US/09314198
; Patent No. 6403310
;  GENERAL INFORMATION:
;    APPLICANT:  Meissner, Paul et al.
;    TITLE OF INVENTION:  Human Inositol Monophosphatase H1
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciencs, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/314,198
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  09/002,072
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  WO PCT/US94/10465
;       FILING DATE:  16-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brookes, A. Anders
;       REGISTRATION NUMBER:  36,373
;       REFERENCE/DOCKET NUMBER:  PF134D1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  301-309-8504
;       TELEFAX:  301-309-8439
;    INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  265 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-314-198-2

  Query Match             10.0%;  Score 173;  DB 4;  Length 265;
  Best Local Similarity   24.8%;  Pred. No. 1.4e-10;
  Matches   67;  Conservative   35;  Mismatches  126;  Indels   42;  Gaps    7;

Qy      91 EVLRKYFRQRVEIIDKEDHSP-VTIADREAEEAMVSVILKSFPTHAIFGEE---NGWRCA 146
           |::||   :   :  |   :  ||  |   |: ::| : : ||:|    ||   :| :|
Db       9 EIIRKALTEEKRVSTKTSAADLVTETDHLVEDLIISELRERFPSHRFIAEEAAASGAKCV 68

Qy     147 ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQT 206
             :   |::|||||| :|:   |     |      :  |||        : |  |:
Db      69 LTHSP-TWIIDPIDGTCNFVHRFPTVAVSIGFAVRQELEFGVIYHCTEERLYTGRRGRGA 127

Qy     207 TLNGQEISVRSCNLLAQAYLYT---------------TSPHLFEADAEDAFIRVRNKVK 250
           |||  : |     |::| : |                     |  | |
Db     128 FCNGQRLRVSGETDLSKALVLTEIGPKRDPATLKLFLSNMGRLLHAKAHG---------- 177

Qy     251 VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
           | : |    |   ||||  |   : ||  :|  :   :|  |||  : |  |  |   |
Db     178 VRVIGSSTLALCHLASGAADAYYQFGLHCWDLAAATVIIREAGGIVIDTSGGPLDLMVC- 236

Qy     311 ESRPTSFNVVAAGD---ARVHKEALDALRW 337
                 ||||     | :  :||  : :
Db     237 -------RVVAASTREMAMLIAQALQTINY 259


RESULT 15
PCT-US94-10465-3
; Sequence 3, Application PC/TUS9410465
;  GENERAL INFORMATION:
;     APPLICANT:  MEISSNER, ET AL.
;     TITLE OF INVENTION:  Human Inositol Monophosphatase H1
;     NUMBER OF SEQUENCES:  3
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;       ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;       STREET:  6 BECKER FARM ROAD
;       CITY:  ROSELAND
;       STATE:  NEW JERSEY
;       COUNTRY:  USA
;       ZIP:  07068
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 INCH DISKETTE
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  WORD PERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/10465
;       FILING DATE:  concurrently
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FERRARO, GREGORY D.
;       REGISTRATION NUMBER:  36,134
;       REFERENCE/DOCKET NUMBER:  325800-127
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;    INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  265 AMINO ACIDS
;       TYPE:  AMINO ACID
;       STRANDEDNESS:
;       TOPOLOGY:  LINEAR
;     MOLECULE TYPE:  PROTEIN
PCT-US94-10465-3

  Query Match             10.0%;  Score 173;  DB 5;  Length 265;
  Best Local Similarity   24.8%;  Pred. No. 1.4e-10;
  Matches   67;  Conservative   35;  Mismatches  126;  Indels   42;  Gaps    7;

Qy      91 EVLRKYFRQRVEIIDKEDHSP-VTIADREAEEAMVSVILKSFPTHAIFGEE---NGWRCA 146
           |::||   :   :  |   :  ||  |   |: ::| : : ||:|    ||   :| :|
Db       9 EIIRKALTEEKRVSTKTSAADLVTETDHLVEDLIISELRERFPSHRFIAEEAAASGAKCV 68

Qy     147 ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQT 206
             :   |::|||||| :|:   |     |      :  |||        : |  |:
Db      69 LTHSP-TWIIDPIDGTCNFVHRFPTVAVSIGFAVRQELEFGVIYHCTEERLYTGRRGRGA 127

Qy     207 TLNGQEISVRSCNLLAQAYLYT---------------TSPHLFEADAEDAFIRVRNKVK 250
           |||  : |     |::| : |                     |  | |
Db     128 FCNGQRLRVSGETDLSKALVLTEIGPKRDPATLKLFLSNMGRLLHAKAHG---------- 177

Qy     251 VPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
           | : |    |   ||||  |   : ||  :|  :   :|  |||  : |  |  |   |
Db     178 VRVIGSSTLALCHLASGAADAYYQFGLHCWDLAAATVIIREAGGIVIDTSGGPLDLMVC- 236

Qy     311 ESRPTSFNVVAAGD---ARVHKEALDALRW 337
                 ||||     | :  :||  : :
Db     237 -------RVVAASTREMAMLIAQALQTINY 259


Search completed: November 17, 2002, 16:43:35
Job time : 48 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 17, 2002, 16:37:33 ; Search time 44 Seconds
(without alignments)
226.022 Million cell updates/sec

Title: US-09-686-522C-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	255.5	14.8	267	4	US-09-118-442-17	Sequence 17, Appl
2	255.5	14.8	267	4	US-09-677-064-17	Sequence 17, Appl
3	211	12.2	277	4	US-09-091-952A-27	Sequence 27, Appl
4	209	12.1	285	4	US-09-091-952A-25	Sequence 25, Appl
5	208	12.0	281	4	US-09-134-001C-3445	Sequence 3445, Ap
6	199.5	11.5	270	1	US-08-347-471-4	Sequence 4, Appli
7	196.5	11.4	288	4	US-09-091-952A-17	Sequence 17, Appl
8	193.5	11.2	277	1	US-08-347-471-2	Sequence 2, Appli
9	193.5	11.2	277	4	US-09-091-952A-26	Sequence 26, Appl
10	193.5	11.2	277	4	US-09-091-952A-28	Sequence 28, Appl
11	173	10.0	265	1	US-08-461-731-2	Sequence 2, Appli
12	173	10.0	265	2	US-09-002-072B-2	Sequence 2, Appli
13	173	10.0	265	4	US-09-314-199-2	Sequence 2, Appli
14	173	10.0	265	4	US-09-314-198-2	Sequence 2, Appli
15	173	10.0	265	5	PCT-US94-10465-3	Sequence 3, Appli
16	104.5	6.0	276	4	US-09-134-001C-5015	Sequence 5015, Ap
17	96.5	5.6	4472	2	US-08-804-227C-2	Sequence 2, Appli
18	90.5	5.2	912	5	PCT-US91-09422-19	Sequence 19, Appl
19	90.5	5.2	1996	2	US-08-804-227C-9	Sequence 9, Appli
20	90.5	5.2	1996	2	US-08-804-198-3	Sequence 3, Appli
21	90	5.2	667	4	US-09-303-064-55	Sequence 55, Appl
22	90	5.2	667	4	US-09-086-503-55	Sequence 55, Appl
23	87	5.0	1792	2	US-08-962-284-4	Sequence 4, Appli
24	86	5.0	912	4	US-08-617-785-2	Sequence 2, Appli
25	86	5.0	912	4	US-09-641-318-2	Sequence 2, Appli
26	85.5	4.9	502	4	US-09-413-814-106	Sequence 106, App
27	85	4.9	283	2	US-08-852-401-4	Sequence 4, Appli
28	84	4.9	3739	3	US-09-320-878-2	Sequence 2, Appli
29	84	4.9	3739	4	US-09-105-537-33	Sequence 33, Appl
30	84	4.9	11877	4	US-09-105-537-6	Sequence 6, Appli
31	83.5	4.8	340	4	US-08-975-762-54	Sequence 54, Appl
32	83.5	4.8	340	4	US-09-295-028-54	Sequence 54, Appl
33	83.5	4.8	340	4	US-09-106-582-54	Sequence 54, Appl
34	83.5	4.8	406	4	US-09-066-046-21	Sequence 21, Appl
35	83.5	4.8	546	4	US-09-352-990-2	Sequence 2, Appli
36	82.5	4.8	367	4	US-09-009-816-2	Sequence 2, Appli
37	82.5	4.8	2910	1	US-08-466-033-183	Sequence 183, App
38	82.5	4.8	2910	2	US-08-444-733-183	Sequence 183, App
39	82.5	4.8	2910	2	US-08-464-134-183	Sequence 183, App
40	82.5	4.8	2910	2	US-08-461-361-183	Sequence 183, App
41	82.5	4.8	2910	2	US-08-485-910-183	Sequence 183, App
42	82.5	4.8	2910	5	PCT-US95-06266-157	Sequence 157, App
43	82.5	4.8	4551	3	US-09-320-878-1	Sequence 1, Appli
44	82.5	4.8	4613	4	US-09-105-537-31	Sequence 31, Appl
45	82	4.7	291	4	US-09-286-691-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-09-118-442-17
; Sequence 17, Application US/09118442B
; Patent No. 6197561
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; APPLICANT: Wang, Xun
; APPLICANT: Bowen, Benjamin A.
; TITLE OF INVENTION: Genes Controlling Phytate Metabolism in
; TITLE OF INVENTION: Plants and Uses Thereof
; FILE REFERENCE: 0706
; CURRENT APPLICATION NUMBER: US/09/118,442B
; CURRENT FILING DATE: 1998-07-17
; EARLIER APPLICATION NUMBER: 60/055,446
; EARLIER FILING DATE: 1997-08-11
; EARLIER APPLICATION NUMBER: 60/055,526
; EARLIER FILING DATE: 1997-08-08
; EARLIER APPLICATION NUMBER: 60/053,944
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Zea mays
US-09-118-442-17

Query Match 14.8%; Score 255.5; DB 4; Length 267;
Best Local Similarity 30.3%; Pred. No. 1.4e-19;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;

```
Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           :: |: : ||  || :|||::|| | |   :  |     ||  |:  |: : : : | ||
Db       1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHKGQVDLVTETDKACEDLIFNHLRKHFP 60

Qy     133 THAIFGEENG---WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
            |   |||       |: : |  |::||:||| :|: | |     : |     | :||:
Db      61 DHKFIGEEESAALGATADLTDDPTWIVDPLDGTTNFVHGFPFVCVSVGLTIGKIPTVGVV 120

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
             ||: | :  | ||   |||  |   | : | :| | | :    :    |      |::
Db     121 FNPIMNELFTAVRGKGAFLNGSPIKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177

Qy     250 KVPLYGCD----CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
             ||       | : ||    :| | :|:  | |   |:|  :   ::: ||| : |
Db     178 NRLLYKIRSIRMCGSLALNMCGVACGRLDLCYEIGFGGPWDVAAGAVILQEAGGLVFDPS 237

Qy     301 G 301
```

```
Db     238 G 238


RESULT 2
US-09-677-064-17
; Sequence 17, Application US/09677064
; Patent No. 6291224
; GENERAL INFORMATION:
;  APPLICANT:  Martino-Catt, Susan J.
;  APPLICANT:  Wang, Hongyu
;  APPLICANT:  Beach, Larry R.
;  TITLE OF INVENTION:  Genes Controlling Phytate Metabolism in
;  TITLE OF INVENTION:  Plants and Uses Thereof
;  FILE REFERENCE:  0706D
;  CURRENT APPLICATION NUMBER: US/09/677,064
;  CURRENT FILING DATE:  2000-09-29
;  PRIOR APPLICATION NUMBER:  60/055,446
;  PRIOR FILING DATE:  1997-08-11
;  PRIOR APPLICATION NUMBER:  60/055,526
;  PRIOR FILING DATE:  1997-08-08
;  PRIOR APPLICATION NUMBER:  60/053,944
;  PRIOR FILING DATE:  1997-07-28
;  PRIOR APPLICATION NUMBER:  09/118,442
;  PRIOR FILING DATE:  1998-07-17
;  NUMBER OF SEQ ID NOS:  31
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 17
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-677-064-17

  Query Match             14.8%;  Score 255.5;  DB 4;  Length 267;
  Best Local Similarity   30.3%;  Pred. No. 1.4e-19;
  Matches   73;  Conservative   41;  Mismatches  112;  Indels   15;  Gaps    5;

Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           :: |: : ||  || :|||::|| | |   :  |     ||  |:  |: : : : | ||
Db       1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHKGQVDLVTETDKACEDLIFNHLRKHFP 60

Qy     133 THAIFGEENG---WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
             |   |||       |: : |  |::||:||| :|: | |     : |     | :||:
Db      61 DHKFIGEEESAALGATADLTDDPTWIVDPLDGTTNFVHGFPFVCVSVGLTIGKIPTVGVV 120

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
             ||: | :  | ||   |||  |   | : | :| | | :    :    |      |::
Db     121 FNPIMNELFTAVRGKGAFLNGSPIKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177

       250 KVPLYGCD----CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
             ||       | : ||    :| | :|:  | |   |:|  :   ::: ||| : |
Db     178 NRLLYKIRSIRMCGSLALNMCGVACGRLDLCYEIGFGGPWDVAAGAVILQEAGGLVFDPS 237

Qy     301 G 301
           |
Db     238 G 238


RESULT 3
US-09-091-952A-27
; Sequence 27, Application US/09091952A
; Patent No. 6458532
;   GENERAL INFORMATION:
;        APPLICANT: Detera-Wadleigh, Sevilla D.
;                   Gershon, Elliot S.
;                   Badner, Judith A.
;                   Goldin, Lynn R.
;                   Berrettini, Wade H.
;                   Yoshikawa, Takeo
;                   Sanders, Alan R.
;                   Esterling, Lisa E.
;        TITLE OF INVENTION: Chromosomal Markers and Diagnostic
;                            Tests for Manic-Depressive Illness
;        NUMBER OF SEQUENCES: 197
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Townsend and Townsend and Crew LLP
;             STREET: Two Embarcadero Center, Eighth Floor
;             CITY: San Francisco
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 94111-3834
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/091,952A
;             FILING DATE: 19-Apr-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 60/029,278
;             FILING DATE: 28-OCT-1996
;             APPLICATION NUMBER: PCT/US97/19381
;             FILING DATE: 28-OCT-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Smith, Timothy L.
;             REGISTRATION NUMBER: 35,367
;             REFERENCE/DOCKET NUMBER: 015280-297100US
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 576-0200
;             TELEFAX: (415) 576-0300
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 27:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 277 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: <Unknown>
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        FEATURE:
;             NAME/KEY:  mat_peptide
;             LOCATION:  1...277
;             OTHER INFORMATION:  bovine IMP
;        SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-091-952A-27

  Query Match             12.2%;  Score 211;  DB 4;  Length 277;
  Best Local Similarity   26.0%;  Pred. No. 1.1e-14;
  Matches   67;  Conservative   52;  Mismatches  115;  Indels   24;  Gaps    8;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSP---VTIADREAEEAMVSVILKSFP 132
           :  :: |   |  ||||:|:  :  : |: |   ||   ||  |:: |: ::: | : :|
Db       6 QECMDYAVTLAGQAGEVVREALKNEMNIMVKS--SPADLVTATDQKVEKMLITSIKEKYP 63

Qy     133 THAIFGEENGWRCAENSA---DFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           :|:  |||:     | |    :  |::|||||| :|: | |     |  : | |   |::
Db      64 SHSFIGEES-VAAGEKSILTDNPTWIIDPIDGTTNFVHGFPFVAVSIGFVVNKKMEFGIV 122

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
             :  : : |  ||    |||:: |     : :: | |    |  :   :    : : :
Db     123 YSCLEDKMYTGRKGKGAFCNGQKLQVSHQEDITKSLLVT---ELGSSRTPETVRIILSNI 179

Qy     250 K----VPLYGCDCYAYA-----LLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300
           :    :|::|      |     |:|:|  |   | |:  :|      ::  ||| : |
Db     180 ERLLCLPIHGIRGVGTAALNMCLVAAGAADAYYEMGIHCWDVAGAGIIVTEAGGVLLDVT 239

Qy     301 G---DKLHWPVTAESRPT 315
           |   | :   | | |  |
Db     240 GGPFDLMSRRVIASSNKT 257


RESULT 4
US-09-091-952A-25
; Sequence 25, Application US/09091952A
; Patent No. 6458532
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 17, 2002, 14:18:58 ; Search time 71 Seconds
(without alignments)
634.348 Million cell updates/sec

Title: US-09-686-522C-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1730	100.0	338	21	AAY44230	Corn extragenic su
2	1294.5	74.8	324	21	AAY44233	Partial wheat extr
3	1067	61.7	295	21	AAY44232	Soybean extragenic
4	1020	59.0	286	23	ABB93299	Herbicidally activ
5	310.5	17.9	260	22	AAG90637	C glutamicum prote
6	310.5	17.9	260	22	AAB79381	Corynebacterium gl
7	310.5	17.9	260	22	AAB79382	Corynebacterium gl
8	286	16.5	265	19	AAW42392	Aquifex pyrophilus
9	279	16.1	264	19	AAW42388	Aquifex VF-5 alkal
10	267.5	15.5	271	21	AAG21330	Arabidopsis thalia
11	260	15.0	361	21	AAG11457	Arabidopsis thalia
12	260	15.0	371	21	AAG11456	Arabidopsis thalia
13	260	15.0	400	21	AAG11455	Arabidopsis thalia
14	255.5	14.8	267	20	AAW97883	Maize myo-inositol
15	253.5	14.7	371	21	AAG39265	Arabidopsis thalia
16	253.5	14.7	400	21	AAG39264	Arabidopsis thalia
17	251.5	14.5	361	21	AAG39266	Arabidopsis thalia
18	244.5	14.1	266	22	AAU67151	Propionibacterium
19	237.5	13.7	267	21	AAY44227	Wheat myo-inositol
20	236.5	13.7	233	21	AAB41258	Human ORFX ORF1022
21	236.5	13.7	233	23	ABP04774	Human ORFX protein
22	224	12.9	339	20	AAW86325	Kidney injury asso
23	221	12.8	168	21	AAY44225	Partial soybean my
24	218	12.6	257	23	ABB47618	Listeria monocytog
25	212	12.3	249	21	AAY44228	Partial barley ext
26	212	12.3	284	22	ABB66131	Drosophila melanog
27	210.5	12.2	273	22	AAG90638	C glutamicum prote
28	210.5	12.2	288	22	ABB66138	Drosophila melanog
29	208.5	12.1	291	22	AAB79387	Corynebacterium gl
30	208.5	12.1	291	22	AAB79388	Corynebacterium gl
31	208	12.0	281	23	ABP38600	Staphylococcus epi
32	207	12.0	254	22	AAB96106	Putative inositol
33	207	12.0	397	21	AAG36298	Arabidopsis thalia
34	204	11.8	264	22	AAU55613	Propionibacterium
35	201	11.6	278	22	ABB71757	Drosophila melanog
36	200	11.6	284	22	ABB66140	Drosophila melanog
37	199.5	11.5	270	15	AAR45333	Rat IMP. Rattus r
38	196.5	11.4	288	19	AAW60676	IMP-18p myo-inosit
39	195.5	11.3	253	22	AAG82667	S. epidermidis ope
40	194.5	11.2	596	22	ABB71755	Drosophila melanog
41	193.5	11.2	277	15	AAR45332	Human brain IMP.
42	193.5	11.2	299	22	ABG16258	Novel human diagno
43	188.5	10.9	325	20	AAY35574	Chlamydia pneumoni
44	187.5	10.8	206	22	AAU67657	Propionibacterium
45	185	10.7	275	23	ABB32472	Staphylococcus aur

priority claim

ALIGNMENTS

RESULT 1
AAY44230
ID AAY44230 standard; Protein; 338 AA.
XX
AC AAY44230;
XX
DT 21-FEB-2000 (first entry)
XX
DE Corn extragenic suppressor protein.
XX
KW Clone cdt2c.pk003.b20; corn extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW animal feed.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT Misc-difference 11
FT /note= "Encoded by GCC"
FT Misc-difference 94
FT /note= "Encoded by ANG"
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR N-PSDB; AAZ29165.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 6; Fig 2; 60pp; English.
XX
CC The present sequence is corn extragenic suppressor protein encoded by the
CC cDNA insert in clone cdt2c.pk003.b20. This belongs to the inositol
CC monophosphate family of proteins. The cDNA clone was isolated from corn
CC developing tassle. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising the corresponding cDNA. This is done to increase
CC levels of free phosphate and decrease phytic acid levels in developing
CC seeds or in the construction of transgenic plants with altered levels of
CC the enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. This is
CC commercially used for animal feed.
XX
SQ Sequence 338 AA;

Query Match 100.0%; Score 1730; DB 21; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.3e-169;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MLSSSSSTHSDTSPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MLSSSSSTHSDTSPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAA 60

Qy      61 GAAGRRAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAE 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GAAGRRAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAE 120

Qy     121 EAMVSVILKSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 EAMVSVILKSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH 180

Qy     181 NGKPVIGVIDQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAED 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 NGKPVIGVIDQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAED 240

Qy     241 AFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300

Qy     301 GDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALRWR 338
          ||||||||||||||||||||||||||||||||||||||
Db     301 GDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALRWR 338
```

RESULT 2
AAY44233
ID AAY44233 standard; Protein; 324 AA.
XX
AC AAY44233;
XX
DT 21-FEB-2000 (first entry)
XX
DE Partial wheat extragenic suppressor protein.
XX
KW Clone wlm0.pk0010.f6; wheat extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW animal feed.
XX
OS Triticum aestivum.
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR N-PSDB; AAZ29168.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 6; Fig 2; 60pp; English.
XX
CC The present sequence is a portion of wheat extragenic suppressor protein
CC encoded by the cDNA insert in clone wlm0.pk0010.f6. This belongs to the
CC inositol monophosphate family of proteins. The cDNA clone was isolated
CC from wheat seedlings. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising the corresponding cDNA. This is done to increase
CC levels of free phosphate and decrease phytic acid levels in developing
CC seeds or in the construction of transgenic plants with altered levels of
CC the enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. This is
CC commercially used for animal feed.
XX
SQ Sequence 324 AA;

Query Match 74.8%; Score 1294.5; DB 21; Length 324;
Best Local Similarity 80.6%; Pred. No. 9.8e-125;
Matches 254; Conservative 24; Mismatches 28; Indels 9; Gaps 3;

```
Qy      32 LRAASPVSSAVLSASGR------QPMSTVRASFAAGAAGRRAAAVGE--LATERLVEVAQ 83
           ||: | ::: ||:||      : | :|||| : | |  || |: : |||| |||
Db      11 LRSPSLLATFSSSAAGRACGIAGRWMGSVRAS-PSEAGGWAVAAAGKEGVDMERLVAVAQ 69

Qy      84 RAADAAGEVLRKYFR RVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGEENGW 143
           |||||||||||||||.| |||||||||||||||||||| ||||||||||:|||||||
Db      70 SAADAAGEVLRKYFRQRFEIIDKEDHSPVTIADREAEEAMTSVILKSFPTHAVFGEENGW 129

Qy     144 RCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDG 203
           |||| |||:|||||||||||||||||||||||||||||:|:|||||||||:||||
Db     130 RCAEKSADYVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVMGIIDQPILRERWVGVDG 189

Qy     204 KQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYGCDCYAYAL 263
           |:|||||||||| ||:| |||||||||||| |||||||||:|||||||||||||||
Db     190 KKTTLNGQEISVRPCNVLEQAYLYTTSPHLFEGDAEDAFIRVRDKVKVPLYGCDCYAYAL 249

Qy     264 LASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPTSFNVVAAG 323
           |||||||:|||||||||||||||||||||||||| |:|||||::|||||||||||||
Db     250 LASGFVDLVVESGLKPYDFLSLVPVIEGAGGSITDWEGNKLHWPVSSESRPTSFNVVAAG 309

Qy     324 DARVHKEALDALRWR 338
           |: || :|| |||||
Db     310 DSHVHGQALAALRWR 324
```

RESULT 3
AAY44232
ID AAY44232 standard; Protein; 295 AA.
XX
AC AAY44232;
XX
DT 21-FEB-2000 (first entry)
XX
DE Soybean extragenic suppressor protein.
XX
KW Clone s12.pk122.p24; clone src3c.pk013.g15; clone sf11n.pk003.g19;
KW soybean extragenic suppressor protein; phytic acid biosynthetic enzyme;
KW transformation; chimeric gene; antibody; free phosphate;

KW phytic acid level; transgenic plant; enzyme detection; animal feed.
XX
OS Glycine max.
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR N-PSDB; AAZ29167.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 6; Fig 2; 60pp; English.
XX
CC The present sequence is a soybean extragenic suppressor protein encoded
CC by a contig assembled from portions of the cDNA inserts in clones
CC s12.pk122.p24, src3c.pk013.g15, and sf11n.pk003.g19. This belongs to the
CC inositol monophosphate family of proteins. The level of expression
CC of this phytic acid biosynthetic enzyme can be altered by transforming a
CC host cell with the chimeric gene comprising the corresponding cDNA. This
CC is done to increase levels of free phosphate and decrease phytic acid
CC levels in developing seeds or in the construction of transgenic plants
CC with altered levels of the enzyme. The phytic acid biosynthetic enzymes
CC are used to raise antibodies, for enzyme detection in cells and cell
CC extracts. This is commercially used for animal feed.
XX
SQ Sequence 295 AA;

Query Match 61.7%; Score 1067; DB 21; Length 295;
Best Local Similarity 63.0%; Pred. No. 2.5e-101;
Matches 201; Conservative 42; Mismatches 44; Indels 32; Gaps 2;

```
Qy      20 SANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGELATERLV 79
           | :| | :   |||| :| |:                              |   |
Db       9 SHSPIPNT-TFRLRAMAPHST------------------------------PLELNRFA 36

Qy      80 EVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           ||  : |||||||:|||||:  ::| | | ||||||| |||||||:|| :||:|||:||
Db      37 EVGNKVADAAGEVIRKYFRKNFDVIHKHDLSPVTIADQSAEEAMVSIILDNFPSHAIYGE 96

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWI 199
           |||||| | :||:|||||||||||||||||:||||:||| || |::|:||||:||||||
Db      97 ENGWRCEEKNADYVWVLDPIDGTKSFITGKPVFGTLVALLQNGTPILGIIDQPVLRERWI 156

Qy     200 GVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYGCDCY 259
           |: ||:|:|||||||| |:|  |:||||||||||||  |||:||||||:|||  |||||||
Db     157 GIAGKRTSLNGQEISTRTCADLSQAYLYTTSPHLFNGDAEEAFIRVRSKVKFQLYGCDCY 216

Qy     260 AYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPTSFNV 319
           |||||:||||||:||||||||||:|:|||||||| |:||:|||| |  :  |  ||||||
Db     217 AYALLSSGFVDLVVESGLKPYDFLALIPVIEGAGGVI DWKGDKLFWEASPLSIATSFNV 276

Qy     320 VAAGDARVHKEALDALRWR 338
           ||||| ::|::|||:|:|:
Db     277 VAAGDKQIHQQALDSLQWK 295
```

RESULT 4
ABB93299
ID ABB93299 standard; Protein; 286 AA.
XX
AC ABB93299;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2510.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP09892.
XX
PR 28-AUG-2001; 2001WO-EP09892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX
PS Claim 5; SEQ ID NO 2510; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
SQ Sequence 286 AA;

Query Match 59.0%; Score 1020; DB 23; Length 286;
Best Local Similarity 69.5%; Pred. No. 1.7e-96;
Matches 182; Conservative 45; Mismatches 35; Indels 0; Gaps 0;

```
Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :|   |    |||:|||:|||||:: :|:||:| |||||||: |||||||:| :: |:||
Db      24 DRFAAVGNALADASGEVIRKYFRKKFDIVDKDDMSPVTIADQMAEEAMVSIIFQNLPSHA 83

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |:||| |||| | |||:||||||||||||||||:||||||||:.|||::|:|||||||:
Db      84 IYGEEKGWRCKEESADYVWVLDPIDGTKSFITGKPVFGTLIALLYKGKPILGLIDQPILK 143

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           |||||::|::| |||::|| |||  |:||||||||||||  :|| |: |||:||||||||
Db     144 ERWIGMNGRRTKLNGEDISTRSCPKLSQAYLYTTSPHLFSEEAEKAYSRVRDKVKVPLYG 203

Qy     256 CDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPT 315
           ||||||||||||||:|:||||||||||:||||||||||:|||| | :  |  :: :  |
Db     204 CDCYAYALLASGFVDLVIESGLKPYDFLALVPVIEGAGGTITDWTGKRFLWEASSSAVAT 263

Qy     316 SFNVVAAGDARVHKEALDALRW 337
           |||||||||: :|::||::| |
Db     264 SFNVVAAGDSDIHQQALESLEW 285
```

RESULT 5
AAG90637
ID AAG90637 standard; Protein; 260 AA.
XX
AC AAG90637;
XX
DT 26-SEP-2001 (first entry)

XX
DE C glutamicum protein fragment SEQ ID NO: 4391.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH65856.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 4391; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 260 AA;

Query Match 17.9%; Score 310.5; DB 22; Length 260;
Best Local Similarity 33.3%; Pred. No. 2.2e-23;
Matches 91; Conservative 42; Mismatches 101; Indels 39; Gaps 11;

```
Qy      81 VAQRAADAAGEV-LRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: |  : | ::    :|:  | | :||: ||   |||:   |  : |  :| ||
Db      10 LALELAELADSITLDRFEASDLEVSSKPDMTPVSDADLATEEALREKIATARPADSILGE 69

Qy     140 ENGWRCAENSADF---VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
           | |        :|    |::||||||::: | |:: |||||| ||||| |||  | |
Db      70 EFG-----GDVEFSGRQWIIDPIDGTKNYVRGVPVWATLIALLDNGKPVAGVISAPALAR 124

Qy     197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTTSPHLFEADAEDAFIRVRNKV- 249
           ||   :|     | ||   :::||   : |  | | :::     | |  | |: : :
Db     125 RWWASEGAGAWRTFNGSSPRKLSVSQVSKLDDASLSFSSLSGWAERDLRDQFVSLTDTTW 184

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG-DKLHWPV 308
           ::  || | ::| |:| | |||  |  :  :|   |  ::  |||  |   | |  |
Db     185 RLRGYG-DFFSYCLVAEGAVDIAAEPEVSLWDLAPLSILVTEAGGKFTSLAGVDGPH--- 240

Qy     309 TAESRPTSFNVVAAGDA-----RVHKEALDALR 336
                       |||      :| | || |:
Db     241 -------------GGDAVATNGILHDETLDRLK 260
```

RESULT 6
AAB79381
ID AAB79381 standard; Protein; 260 AA.
XX
AC AAB79381;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:278.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100844-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061975/07.
DR N-PSDB; AAF71498.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 20; Page 537-538; 1246pp; English.
XX
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar

```
CC   metabolism and oxidative phosphorylation (SMP) proteins given in
CC   AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC   energy production. The C. glutamicum SMP gene can be used in vectors
CC   (II) for expression in host cells and production or modulation of
CC   production of fine chemicals, such as, an organic acid, a proteinogenic
CC   or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC   a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC   acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC   cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC   (III) encoded by them are used for diagnosing the presence or activity of
CC   Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC   containing them are used to map genomes of organisms related to
CC   C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC   in evolutionary studies, in determining SMP protein regions required
CC   for function, in modulating SMP protein activity, in modulating the
CC   metabolism of sugars, and in modulating high-energy molecule production
CC   in a cell (i.e. ATP, NADPH).
XX
SQ   Sequence   260 AA;

  Query Match             17.9%;  Score 310.5;  DB 22;  Length 260;
  Best Local Similarity   33.3%;  Pred. No. 2.2e-23;
  Matches   91;  Conservative   42;  Mismatches  101;  Indels   39;  Gaps   11;

Qy      81 VAQRAADAAGEV-LRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: |  : | ::    :|:  | | :||: ||   |||:   |  : |  :| ||
Db      10 LALELAELADSITLDRFEASDLEVSSKPDMTPVSDADLATEEALREKIATARPADSILGE 69

Qy     140 ENGWRCAENSADF---VWVLDPIDGTKSFITGKPLFGTLIALLHNGKP GVIDQPILRE 196
           | |        :|    |::|||||||::: | |:: |||||| |||| ||| | |
Db      70 EFG-----GDVEFSGRQWIIDPIDGTKNYVRGVPVWATLIALLDNGKPVAGVISAPALAR 124

Qy     197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTTSPHLFEADAEDAFIRVRNKV- 249
           ||   :|     | ||   :::||   : |  | | :::     | |  | |: : :
Db     125 RWWASEGAGAWRTFNGSSPRKLSVSQVSKLDDASLSFSSLSGWAERDLRDQFVSLTDTTW 184

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLV VIEGAGGSITDWRG-DKLHWPV 308
           ::  || | ::| |:| | |||  |  :  :|   |  ::  |||  |   | | |
Db     185 RLRGYG-DFFSYCLVAEGAVDIAAEPEVSLWDLAPLSILVTEAGGKFTSLAGVDGPH--- 240

Qy     309 TAESRPTSFNVVAAGDA-----RVHKEALDALR 336
                     |||      :| | || |:
Db     241 ------------GGDAVATNGILHDETLDRLK 260


RESULT 7
AAB79382
ID   AAB79382 standard; Protein; 260 AA.
XX
AC   AAB79382;
XX
DT   30-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum SMP protein sequence SEQ ID NO:280.
XX
KW   Corynebacterium glutamicum; carbon metabolism and energy production;
KW   SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW   fine chemical production; organic acid; proteinogenic amino acid;
KW   nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW   nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW   carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW   diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100844-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00943.
XX
PR   25-JUN-1999;   99US-0141031.
PR   08-JUL-1999;   99DE-1031412.
PR   08-JUL-1999;   99DE-1031413.
PR   08-JUL-1999;   99DE-1031419.
PR   08-JUL-1999;   99DE-1031420.
PR   08-JUL-1999;   99DE-1031424.
PR   08-JUL-1999;   99DE-1031428.
PR   08-JUL-1999;   99DE-1031431.
PR   08-JUL-1999;   99DE-1031433.
PR   08-JUL-1999;   99DE-1031434.
PR   08-JUL-1999;   99DE-1031510.
PR   08-JUL-1999;   99DE-1031562.
PR   08-JUL-1999;   99DE-1031634.
PR   09-JUL-1999;   99DE-1032180.
PR   09-JUL-1999;   99DE-1032227.
PR   09-JUL-1999;   99DE-1032230.
PR   09-JUL-1999;   99US-0143208.
PR   14-JUL-1999;   99DE-1032924.
PR   14-JUL-1999;   99DE-1032973.
PR   14-JUL-1999;   99DE-1033005.
PR   27-AUG-1999;   99DE-1040765.
PR   31-AUG-1999;   99US-0151572.
PR   03-SEP-1999;   99DE-1042076.
PR   03-SEP-1999;   99DE-1042079.
PR   03-SEP-1999;   99DE-1042086.
PR   03-SEP-1999;   99DE-1042087.
PR   03-SEP-1999;   99DE-1042088.
PR   03-SEP-1999;   99DE-1042095.
PR   03-SEP-1999;   99DE-1042123.
PR   03-SEP-1999;   99DE-1042125.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-061975/07.
DR   N-PSDB; AAF71499.
XX
PT   New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT   metabolism and oxidative phosphorylation protein for production or
PT   modulation of production of fine chemicals e.g. amino acids,
PT   carbohydrates or enzymes -
XX
PS   Claim 20; Page 540; 1246pp; English.
XX
CC   AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC   metabolism and oxidative phosphorylation (SMP) proteins given in
CC   AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC   energy production. The C. glutamicum SMP gene can be used in vectors
CC   (II) for expression in host cells and production or modulation of
CC   production of fine chemicals, such as, an organic acid, a proteinogenic
CC   or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC   a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC   acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC   cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC   (III) encoded by them are used for diagnosing the presence or activity of
CC   Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC   containing them are used to map genomes of organisms related to
CC   C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC   in evolutionary studies, in determining SMP protein regions required
CC   for function, in modulating SMP protein activity, in modulating the
CC   metabolism of sugars, and in modulating high-energy molecule production
CC   in a cell (i.e. ATP, NADPH).
XX
SQ   Sequence   260 AA;

  Query Match             17.9%;  Score 310.5;  DB 22;  Length 260;
  Best Local Similarity   33.3%;  Pred. No. 2.2e-23;
  Matches   91;  Conservative   42;  Mismatches  101;  Indels   39;  Gaps   11;

Qy      81 VAQRAADAAGEV-LRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: |  : | ::    :|:  | | :||: ||   |||:   |  : |  :| ||
Db      10 LALELAELADSITLDRFEASDLEVSSKPDMTPVSDADLATEEALREKIATARPADSILGE 69

Qy     140 ENGWRCAENSADF---VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
```

```
          | |        :|    |::||||||::: | |:: |||||| ||||| |||  | |
Db     70 EFG-----GDVEFSGRQWIIDPIDGTKNYVRGVPVWATLIALLDNGKPVAGVISAPALAR 124

Qy    197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTTSPHLFEADAEDAFIRVRNKV- 249
          ||   :|     | ||   :::||   : |  | | :::     | |  | |: : :
Db    125 RWWASEGAGAWRTFNGSSPRKLSVSQVSKLDDASLSFSSLSGWAERDLRDQFVSLTDTTW 184

Qy    250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG-DKLHWPV 308
          ::  || | ::| |:| | |||  |  :  :|   |  ::  |||  |   | |  |
Db    185 RLRGYG-DFFSYCLVAEGAVDIAAEPEVSLWDLAPLSILVTEAGGKFTSLAGVDGPH--- 240

Qy    309 TAESRPTSFNVVAAGDA-----RVHKEALDALR 336
                     |||      :| | || |:
Db    241 -------------GGDAVATNGILHDETLDRLK 260


RESULT 8
AAW42392
ID   AAW42392 standard; Protein; 265 AA.
XX
AC   AAW42392;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Aquifex pyrophilus phosphatase (28ph1).
XX
KW   Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW   food; detergent; baking.
XX
OS   Aquifex pyrophilus.
XX
PN   WO9748416-A1.
XX
PD   24-DEC-1997.
XX
PF   19-JUN-1997;   97WO-US10784.
XX
PR   19-JUN-1996;   96US-0033752.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Bylina E,  Lee E,  Mathur EJ;
XX
DR   WPI; 1998-062851/06.
DR   N-PSDB; AAV03317.
XX
PT   Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT   detergent, and baking industries
XX
PS   Claim 11; Page 95; 128pp; English.
XX
CC   This protein comprises a thermostable phosphatase, designated
CC   28ph1, of Aquifex pyrophilus.  The invention relates to claimed
CC   polynucleotides (see AAV03301-20) coding for claimed thermostable
CC   phosphatases (see AAW42380-95).  Vector and host cells are used to
CC   produce the enzymes, which can be used in a claimed method to
CC   hydrolyse phosphate bonds.  They can also be used in enzyme
CC   labelling processes, in certain recombinant DNA techniques, in
CC   ELISA immunoassays, in enzyme linked gene probes, in research
CC   applications for removing 5' phosphates in polynucleotides prior to
CC   end labelling, and in the pharmaceutical, food, detergent, and
CC   baking industries.
XX
SQ   Sequence   265 AA;

  Query Match             16.5%;  Score 286;  DB 19;  Length 265;
  Best Local Similarity   29.8%;  Pred. No. 7.7e-21;
  Matches   78;  Conservative   53;  Mismatches  105;  Indels   26;  Gaps    8;

Qy      76 ERLVEVAQRAADAAGEVLRKYF-RQRVE-IIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           |:|:|||: || | |:||:: | : ::| | :| :   |:  |: :|| :  :||| ||
Db       5 EKLLEVAKMAALAGGQVLKENFGKIKLENIEEKGEKDFVSYVDKTSEERIKELILKFFPD 64
```

```
Qy    134 HAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPI 193
          | : ||| |    |  : : | :||:||||::| | |:|   : |:   :|::| :  |
Db     65 HEVVGEERGKEGKE--SPYKWFIDPLDGTKNYIKGFPIFAVSVGLVKENEPIVGAVYLPY 122

Qy    194 LRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRVR 246
                :    |:    ||: |||:    |  | :   |     |         :: |  |
Db    123 FDTLYWASKGRGAYKNGERISVKERGELKHAAVVYGFPSRSRRDISLYLNVFKEVFYEV- 181

Qy    247 NKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHW 306
            |: |  |       :|| |  | ::|  :||:|  : : ::: |||  |  :||
Db    182 GSVRRP--GAAAVDICMLAEGIFDGMMEFEMKPWDITAGLVILKEAGGFYT-LKGD---- 234

Qy    307 PVTAESRPTSFNVVAAGDARVH 328
                |   : : ||:  :|
Db    235 -------PFGISDIIAGNRMLH 249


RESULT 9
AAW42388
ID   AAW42388 standard; Protein; 264 AA.
XX
AC   AAW42388;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Aquifex VF-5 alkaline phosphatase (34A1A).
XX
KW   Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW   food; detergent; baking.
XX
OS   Aquifex sp. strain VF-5.
XX
PN   WO9748416-A1.
XX
PD   24-DEC-1997.
XX
PF   19-JUN-1997;   97WO-US10784.
XX
PR   19-JUN-1996;   96US-0033752.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI   Bylina E,  Lee E,  Mathur EJ;
XX
DR   WPI; 1998-062851/06.
DR   N-PSDB; AAV03309.
XX
PT   Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT   detergent, and baking industries
XX
PS   Claim 10; Fig 9; 128pp; English.
XX
CC   This protein comprises a mature thermostable alkaline phosphatase
CC   (34A1A) of Aquifex VF5, a marine bacterium that grows optimally
CC   at 85-90 degC and pH 6.8.  Its amino acid sequence, deduced from
CC   an isolated polynucleotide (see AAV03309), shows 57% identity to
CC   Escherichia coli suppressor protein suhB.  The invention relates to
CC   claimed polynucleotides (see AAV03301-09 and AAV03310-20) coding for
CC   claimed thermostable phosphatases (see AAW42380-95).  Vector and host
CC   cells are used to produce the enzymes, which can be used in a
CC   claimed method to hydrolyse phosphate bonds.  They can also be used
CC   in enzyme labelling processes, in recombinant DNA techniques, in
CC   ELISA, in enzyme linked gene probes, in research applications for
CC   removing 5' phosphates in polynucleotides prior to end labelling,
CC   and in the pharmaceutical, food, detergent, and baking industries.
XX
SQ   Sequence   264 AA;

  Query Match             16.1%;  Score 279;  DB 19;  Length 264;
  Best Local Similarity   28.5%;  Pred. No. 4e-20;
  Matches   75;  Conservative   54;  Mismatches  106;  Indels   28;  Gaps    7;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
```

```
          :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  |||| ||
Db      5 KKYLEVAKIAALAGGQVLKENFGKVKKENIEEKGEKDFVSYVDKTSEERIKEVILKFFPD 64

Qy    134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
          | : ||| |   || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db     65 HEVVGEEMG---AEGSGSEYRWFIDPLDGTKNYINGFPIFAVSVGLVKGEEPIVGAVYLP 121

Qy    193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRV 245
             : : |  |    :||: | |:    |  | :    |     |         :| |  |
Db    122 YFDKLYWGAKGLGAYVNGKRIKVKDNESLKHAGVVYGFPSRSRRDISIYLNIFKDVFYEV 181

Qy    246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
             :: |  |       ::| |  | ::|  :||:|  : : ::: |||
Db    182 -GSMRRP--GAAAVDLCMVAEGIFDGMMEFEMKPWDITAGLVILKEAGG----------- 227

Qy    306 WPVTAESRPTSFNVVAAGDARVH 328
             |    |   : : ||:  :|
Db    228 -VYTLVGEPFGVSDIIAGNKALH 249
```

ESULT 10
AAG21330
ID AAG21330 standard; Protein; 271 AA.
XX
AC AAG21330;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 23850.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.5%; Score 267.5; DB 21; Length 271;
Best Local Similarity 30.1%; Pred. No. 6.5e-19;
Matches 81; Conservative 50; Mismatches 121; Indels 17; Gaps 7;

```
Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :  |  ||  ||:::|| | :   :  |     ||  |:  || : : : : || |
Db       8 DQFLAAAIDAAKKAGQIIRKGFYETKHVEHKSQVDLVTETDKGCEELVFNHLKQLFPNHK 67

Qy     136 IFGEEN--GWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPI 193
             |||    :   | : :  |::||:||| :|: | |     | |     ||:||:  ||
Db      68 FIGEETAAAFGVTELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIGKVPVVGVVYNPI 127

Qy     194 LRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT-TSPHLFEADAEDAFIRVRN---KV 249
           : | : || ||   |||: | | : : |  | | |       :|  :|   |: :   ||
Db     128 MEELFTGVQGKGAFLNGKRIKVSAQSELLTALLVTEAGTKRDKATLDDTTNRINSLLTKV 187

Qy     250 K-VPLYGCDCYAYALLASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWRGDKLHWP 307
           : : : |       :| | |||  | |   |:|  : : ::: ||| | | |  | |
Db     188 RSLRMSGSCALDLCGVACGRVDIFYELGFGGPWDIAAGIVIVKEAGGLIFDPSGKDL--D 245

Qy     308 VTAESRPTSFNVVAAGDARVHKEALDALR 336
           :|::        :|| :| : :   :|||
Db     246 ITSQR-------IAASNASLKELFAEALR 267
```

RESULT 11
AAG11457
ID AAG11457 standard; Protein; 361 AA.
XX
AC AAG11457;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10179.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 260; DB 21; Length 361;
Best Local Similarity 26.9%; Pred. No. 5.9e-18;
Matches 92; Conservative 54; Mismatches 152; Indels 44; Gaps 10;

```
Qy      13 SPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGE 72
           :|| |  : |   | |  |: : |  |:  |    :  :| |          |  |   |
Db      17 NPFSG-RTVNRTFRYRCTRILSNSFKSTTRLQT--KAVLSEVSDQTRYPRIG--AKTTGT 71

Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           ::   |:|| : ||    ||: :   :   |  |     ||  |: :| |:: |: |:|
Db      72 ISPAHLLEVVELAAKTGAEVVMEAVNKPRNITYKGLSDLVTDTDKASEAAILEVVKKNFS 131

Qy     133 THAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKP----VIGV 188
           | | ||| |    ::|:|::| :||:||| :|  | | |   : :|: | |    |:
Db     132 DHLILGEEGG-IIGDSSSDYLWCIDPLDGTTNFAHGYPSFAVSVGVLYRGNPAAASVVEF 190

Qy     189 IDQPILR--ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT----------TSPHLFE 235
           :  |: |    :    |     |||:| |   : : :| | |           |:  ||:
Db     191 VGGPMCRNTRTFSATAGGGALCNGQKIHVSKTDAVERALLITGFGYEHDDAWSTNMELFK 250

Qy     236 ADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGS 295
                 |  |   |:    |        :| |  :   |  |||:|  : | ::| |||:
Db     251 -----EFTDVSRGVR--RLGAAAVDMCHVALGIAESYWEYRLKPWDMAAGVLIVEEAGGA 303

Qy     296 ITDWRGDK-------------LHWPVTAESRPTSFNVVAAG 323
           :|   | |              ||  :     | : |: : |
Db     304 VTRMDGGKFSVFDRSVLVSNGVLHPKLLERIAPATENLKSKG 345
```

RESULT 12
AAG11456
ID AAG11456 standard; Protein; 371 AA.
XX
AC AAG11456;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10178.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 260; DB 21; Length 371;
Best Local Similarity 26.9%; Pred. No. 6.2e-18;
Matches 92; Conservative 54; Mismatches 152; Indels 44; Gaps 10;

```
Qy      13 SPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGE 72
           :|| |  : |   | |  |: : |  |:  |    :  :| |          |  |   |
Db      27 NPFSG-RTVNRTFRYRCTRILSNSFKSTTRLQT--KAVLSEVSDQTRYPRIG--AKTTGT 81

Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           ::   |:|| : ||    ||: :   :   |  |      ||  |: :| |:: |: |:|
Db      82 ISPAHLLEVVELAAKTGAEVVMEAVNKPRNITYKGLSDLVTDTDKASEAAILEVVKKNFS 141

Qy     133 THAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKP----VIGV 188
           | | ||| |    ::|:|::| :||:||| :|  | | |   : :|: | |    |:
Db     142 DHLILGEEGG-IIGDSSSDYLWCIDPLDGTTNFAHGYPSFAVSVGVLYRGNPAAASVVEF 200

Qy     189 IDQPILR--ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT----------TSPHLFE 235
           :  |: |    :    |     |||:| |   : : :| | |           |:  ||:
Db     201 VGGPMCRNTRTFSATAGGGALCNGQKIHVSKTDAVERALLITGFGYEHDDAWSTNMELFK 260

Qy     236 ADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGS 295
               |  |   |:    |         :| |  :   |  |||:|  : | ::| |||:
Db     261 -----EFTDVSRGVR--RLGAAAVDMCHVALGIAESYWEYRLKPWDMAAGVLIVEEAGGA 313

Qy     296 ITDWRGDK-------------LHWPVTAESRPTSFNVVAAG 323
           :|   | |               ||  :     | : |: : |
Db     314 VTRMDGGKFSVFDRSVLVSNGVLHPKLLERIAPATENLKSKG 355
```

RESULT 13
AAG11455
ID AAG11455 standard; Protein; 400 AA.
XX
AC AAG11455;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10177.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX

PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 260; DB 21; Length 400;
Best Local Similarity 26.9%; Pred. No. 6.9e-18;
Matches 92; Conservative 54; Mismatches 152; Indels 44; Gaps 10;

```
Qy      13 SPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGE 72
           :|| |  : |   | |  |: : |  |:  |    :  :| |          |  |   |
Db      56 NPFSG-RTVNRTFRYRCTRILSNSFKSTTRLQT--KAVLSEVSDQTRYPRIG--AKTTGT 110

Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSP. ADREAEEAMVSVILKSFP 132
           ::   |:|| : ||    ||: :   :   |  |     ||  |: :| |:: |: |:|
Db     111 ISPAHLLEVVELAAKTGAEVVMEAVNKPRNITYKGLSDLVTDTDKASEAAILEVVKKNFS 170

Qy     133 THAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKP----VIGV 188
           | | ||| |    ::|:|::| :|||:||| :|  | | |   : :|: | |    |:
Db     171 DHLILGEEGG IIGDSSSDYLWCIDPLDGTTNFAHGYPSFAVSVGVLYRGNPAAASVVEF 229

Qy     189 IDQPILR--ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT----------TSPHLFE 235
           :  |: |   :    |     |||:| |   : : :| | |           |:  ||:
Db     230 VGGPMCRNTRTFSATAGGGALCNGQKIHVSKTDAVERALLITGFGYEHDDAWSTNMELFK 289

Qy     236 ADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGS 295
               |  |   |:    |        :| |  :   |  |||:|  : | ::| |||:
Db     290 -----EFTDVSRGVR--RLGAAAVDMCHVALGIAESYWEYRLKPWDMAAGVLIVEEAGGA 342

Qy     296 ITDWRGDK-------------LHWPVTAESRPTSFNVVAAG 323
           :|   | |              ||  :     | : |: : |
Db     343 VTRMDGGKFSVFDRSVLVSNGVLHPKLLERIAPATENLKSKG 384
```

RESULT 14
AAW97883
ID AAW97883 standard; Protein; 267 AA.
XX
AC AAW97883;

XX
DT 07-JUN-1999 (first entry)
XX
DE Maize myo-inositol monophosphatase-3.
XX
KW Myo-inositol monophosphatase-3; maize; corn; phytate;
KW phytic acid; transgenic plant; animal nutrition; feedstuff; food.
XX
OS Zea mays.
XX
PN WO9905298-A1.
XX
PD 04-FEB-1999.
XX
PF 17-JUL-1998; 98WO-U 14657.
XX
PR 18-MAY-1998; 98US-0085852.
PR 22-JUL-1997; 97US-0053371.
PR 28-JUL-1997; 97US-0053944.
PR 08-AUG-1997; 97US-0055526.
PR 11-AUG-1997; 97US-0055446.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Beach LR, Bowen BA, Martino-Catt SJ, Wang H, Wang X;
XX
DR WPI; 1999-142948/12.
DR N-PSDB; AAX24412.
XX
PT New polynucleotides controlling phytate metabolism in plants -
PT useful for improving the nutritional content of plants, by enhancing
PT levels of non-phytate phosphorus, and reducing phytate levels
XX
PS Claim 1a; Page 81-82; 86pp; English.
XX
CC This is the amino acid sequence of maize myo-inositol
CC monophosphatase-3, an enzyme involved in the metabolism of
CC phytate. cDNA (see AAX24410) encoding the enzyme was isolated from
CC a maize immature ear cDNA library. Polynucleotides (see AAX24400,
CC AAX24403, AAX24407 and AAX24410-12) encoding maize
CC phosphatidylinositol-3-kinase (see AAW97880), myo-inositol
CC 1,3,4-triphosphate 5/6-kinase (see AAW97881), myo-inositol
CC 1-phosphate synthase and myo-inositol monophosphatase-3, all
CC enzymes involved in phytate metabolism, are claimed. The
CC invention relates to the use of such genes to reduce the
CC levels of phytate, and/or increase the levels of non-phytate
CC phosphorus, in plants used for food or feed. The genes are
CC especially used to improve the nutritional content of plants such
CC as corn and soybean. Transgenic plants, and seed produced by them,
CC are claimed.
XX
SQ Sequence 267 AA;

Query Match 14.8%; Score 255.5; DB 20; Length 267;
Best Local Similarity 30.3%; Pred. No. 1.1e-17;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;

```
Qy     73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
          :: |: : ||  || :|||::|| | |   :  |     ||  |:  |: : : : | ||
Db      1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHKGQVDLVTETDKACEDLIFNHLRKHFP 60

Qy    133 THAIFGEENG---WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           |   |||        |: : |  |::||:||| :|: | |     : |     | :||:
Db     61 DHKFIGEEESAALGATADLTDDPTWIVDPLDGTTNFVHGFPFVCVSVGLTIGKIPTVGVV 120

Qy    190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           ||: | :  | ||   |||  |   | : | :| | | :    :    |      |::
Db    121 FNPIMNELFTAVRGKGAFLNGSPIKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177

Qy    250 KVPLYGCD----CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
             ||       | : ||    :| | :|:  | |   |:|  :   ::: ||| : |
Db    178 NRLLYKIRSIRMCGSLALNMCGVACGRLDLCYEIGFGGPWDVAAGAVILQEAGGLVFDPS 237
```

```
Qy     301 G 301
           |
Db     238 G 238


RESULT 15
AAG39265
ID   AAG39265 standard; Protein; 371 AA.
XX
AC   AAG39265;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 48558.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
RN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
```

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.7%; Score 253.5; DB 21; Length 371;
Best Local Similarity 25.8%; Pred. No. 2.9e-17;
Matches 93; Conservative 55; Mismatches 165; Indels 47; Gaps 10;

```
Qy       1 MLSSSSSTHSDTSPFPGLASANP---NPRSRLLRLRAASPVSSAVLSASGRQP---MSTV 54
           : | : |      |   |   ||      :|  | |    :|::  | :  |    :! |
Db       6 IFSGNMSLRISHLPRSSLPLQNPISGRTVNRTFPYRCTRILSNSFKSTTRLQTKAVLSEV 65

Qy      55 RASFAAGAAGRRAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTI 114
                 |  |   | ::   |:|| : ||    ||: :   :   |  |     ||
Db      66 SDQTRYPRIG--AKTTGTISPAHLLEVVELAAKTGAEVVMEAVNKPRNITYKGLSDLVTD 123

Qy     115 ADREAEEAMVSVILKSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGT 174
           |: :| |:: |: |:|  | | ||| |    ::|:|::| :||:||| :|  | | |
Db     124 TDKASEAAILEVVKKNFSDHLILGEEGG-IIGDSSSDYLWCIDPLDGTTNFAHGYPSFAV 182

Qy     175 LIALLHNGKP----VIGVIDQPIL--RERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT 228
           : :|: | |    |:  :  |:      :    |     |||:| |   : : :| | |
Db     183 SVGVLYRGNPAAASVVEFVGGPMCWNTRTFSATAGGGALCNGQKIHVSKTDAVERALLIT 242

Qy     229 -----------TSPHLFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGL 277
                     |:  ||:      |  |   |:    |        :| |  :   | |
Db     243 GFGYEHDDAWSTNMELFK-----EFTDVSRGVR--RLGAAAVDMCHVALGIAESYWEYRL 295

Qy     278 KPYDFLSLVPVIEGAGGSITDWRGDK-------------LHWPVTAESRPTSFNVVAAG 323
           ||:|  : | ::| |||::|   | |              ||  :     | : |: : |
Db     296 KPWDMAAGVLIVEEAGGAVTRMDGGKFSVFDRSVLVSNGVLHPKLLERIAPATENLKSKG 355
```

Search completed: November 17, 2002, 16:39:13
Job time : 99 secs

OM protein - protein search, using sw model

Run on: November 17, 2002, 15:16:18 ; Search time 85 Seconds
(without alignments)
819.340 Million cell updates/sec

Title: US-09-686-522C-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1020	59.0	286	10	Q9T021	Q9t021 arabidopsis
2	551	31.8	259	16	Q92M31	Q92m31 rhizobium m
3	525	30.3	297	16	Q8U945	Q8u945 agrobacteri
4	483	27.9	308	16	Q930E4	Q930e4 rhizobium m
5	477	27.6	278	16	Q9A643	Q9a643 caulobacter
6	463.5	26.8	272	16	Q8YDX6	Q8ydx6 brucella me
7	460	26.6	263	2	Q9KWB5	Q9kwb5 agrobacteri
8	449	26.0	262	16	Q98H43	Q98h43 rhizobium l
9	396.5	22.9	267	16	P73806	P73806 synechocyst
10	317	18.3	239	16	Q988Q3	Q988q3 rhizobium l
11	317	18.3	273	2	Q53743	Q53743 streptomyce
12	302.5	17.5	268	2	Q8RJW8	Q8rjw8 streptomyce
13	293	16.9	267	16	Q8ZCR8	Q8zcr8 yersinia pe
14	283	16.4	255	16	O32889	O32889 mycobacteri
15	278.5	16.1	266	16	Q9K4B1	Q9k4b1 streptomyce
16	273.5	15.8	260	16	P95189	P95189 mycobacteri
17	265.5	15.3	271	10	Q9M8S8	Q9m8s8 arabidopsis
18	258.5	14.9	371	10	Q94F00	Q94f00 arabidopsis
19	258	14.9	267	16	Q8X2E6	Q8x2e6 escherichia
20	257	14.9	267	16	Q8Z4M9	Q8z4m9 salmonella
21	254.5	14.7	270	16	Q8YT10	Q8yt10 anabaena sp
22	249	14.4	270	16	Q8Y084	Q8y084 ralstonia s
23	245	14.2	266	16	Q8YCG2	Q8ycg2 brucella me
24	244.5	14.1	276	16	Q8U6M0	Q8u6m0 agrobacteri
25	244	14.1	263	16	Q8YIW3	Q8yiw3 brucella me
26	240	13.9	266	16	Q8U9I0	Q8u9i0 agrobacteri
27	239.5	13.8	266	17	Q9HQ12	Q9hq12 halobacteri
28	237	13.7	272	16	Q92RQ1	Q92rq1 rhizobium m
29	236.5	13.7	268	16	Q8UHI8	Q8uhi8 agrobacteri
30	231.5	13.4	353	10	Q9SA15	Q9sa15 arabidopsis
31	230	13.3	221	2	P71487	P71487 methylomicr
32	230	13.3	287	16	P74542	P74542 synechocyst
33	229.5	13.3	275	16	Q8YGM7	Q8ygm7 brucella me
34	229	13.2	264	16	Q98D39	Q98d39 rhizobium l
35	222	12.8	295	16	Q8U602	Q8u602 agrobacteri
36	221	12.8	266	16	Q92WJ1	Q92wj1 rhizobium m
37	221	12.8	272	2	O30546	O30546 agrobacteri
38	221	12.8	280	16	Q8UCQ5	Q8ucq5 agrobacteri
39	220	12.7	275	16	Q98NP3	Q98np3 rhizobium l
40	220	12.7	294	16	Q92X03	Q92x03 rhizobium m
41	219	12.7	290	11	Q91UZ5	Q91uz5 mus musculu
42	218	12.6	257	16	Q8Y852	Q8y852 listeria mo
43	218	12.6	304	16	Q8UEA3	Q8uea3 agrobacteri
44	218	12.6	319	5	Q23493	Q23493 caenorhabdi
45	215	12.4	255	2	Q9R6D4	Q9r6d4 agrobacteri

ALIGNMENTS

RESULT 1
Q9T021
ID Q9T021 PRELIMINARY; PRT; 286 AA.
AC Q9T021;
DT 01-MAY-2000 (TrEMBLre . 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Inositol monophosphatase-like protein.
GN T22F8.20 OR AT4G39120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Grueneisen A., Wambutt R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL050351; CAB43627.1; -.
DR EMBL; AL161594; CAB80575.1; -.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
SQ SEQUENCE 286 AA; 31635 MW; 5C2BB2BA90FD10BB CRC64;

```
  Query Match             59.0%;  Score 1020;  DB 10;  Length 286;
  Best Local Similarity   69.5%;  Pred. No. 2.3e-74;
  Matches  182;  Conservative   45;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :|   |    |||:|||:||||:: :|:||:| ||||||| |||||||:| :: |:||
Db      24 DRFAAVGNALADASGEVIRKYFRKKFDIVDKDDMSPVTIADQMAEEAMVSIIFQNLPSHA 83

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |:||| |||| | |||:||||||||||||||||:|||||||: |||::|:||||||:
Db      84 IYGEEKGWRCKEESADYVWVLDPIDGTKSFITGKPVFGTLIALLYKGKPILGLIDQPILK 143

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           |||||::|::| |||::|| |||  |:||||||||||||  :|| |: |||:||||||||
Db     144 ERWIGMNGRRTKLNGEDISTRSCPKLSQAYLYTTSPHLFSEEAEKAYSRVRDKVKVPLYG 203

Qy     256 CDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPT 315
           |||||||||||||||:|:||||||||||:||||||||||:|||| | :  |  :: :  |
Db     204 CDCYAYALLASGFVDLVIESGLKPYDFLALVPVIEGAGGTITDWTGKRFLWEASSSAVAT 263

  /    316 SFNVVAAGDARVHKEALDALRW 337
           |||||||||: :|::||::| |
Db     264 SFNVVAAGDSDIHQQALESLEW 285


RESULT 2
Q92M31
ID   Q92M31      PRELIMINARY;      PRT;   259 AA.
AC   Q92M31;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative monophosphatase protein.
GN   R02825 OR SMC04042.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR   EMBL; AL591791; CAC47404.1; -.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00094; C5_MTASE_1; UNKNOWN_1.
DR   PROSITE; PS00629; IMP_1; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   259 AA;  28162 MW;  2062305CB4B3C871 CRC64;

  Query Match             31.8%;  Score 551;  DB 16;  Length 259;
  Best Local Similarity   44.0%;  Pred. No. 1.5e-36;
  Matches  117;  Conservative   48;  Mismatches   73;  Indels   28;  Gaps    7;

Qy      84 RAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
           | ||||       ||    :::| :    ||| ||| || :: ::|  ||| | | |||:
Db      10 RLADAAKAETMPRFRVGTSVLNKLEGGFDPVTEADRSAEASIRALIESSFPEHGILGEEH 69

Qy     142 GWRCAEN---SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
           |     |     : |||:||||||::||:| |::|||| |  ||| |:|::|||   ||:
Db      70 G-----NIGLDRELVWVIDPIDGTRAFISGLPVWGTLIGLYRNGKAVMGLMDQPFTGERF 124

Qy     199 IGVDGKQTTLNGQE----ISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
             ||::    | :    :: | |: |: | |:|||||:  : :: |  :: ||::  |
Db     125 F-ADGEKALYRGPDGERVLATRPCHALSDAVLFTTSPHLYTGELKERFEALQEKVRLFRY 183

Qy     255 GCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRP 314
           ||||||:||||:| ||:||| ||||||   |:|:|| ||| ||||:|            |
Db     184 GCDCYAFALLAAGHVDLVVECGLKPYDVGGLIPLIEQAGGIITDWQGG----------P 232

Qy     315 TSF--NVVAAGDARVHKEALDALRWR 338
                 ::|||   :| :||:||: |
Db     233 AEMGGEIIAAGSRELHAQALEALKGR 258


RESULT 3
Q8U945
ID   Q8U945      PRELIMINARY;      PRT;   297 AA.
AC   Q8U945;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Inositol monophosphatase family protein.
GN   ATU3885 OR AGR_L_1926.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009319; AAL44693.1; ALT_INIT.
DR   EMBL; AE008295; AAK89535.1; -.
KW   Complete proteome.
SQ   SEQUENCE   297 AA;  33043 MW;  83E08201E4508B1E CRC64;

  Query Match             30.3%;  Score 525;  DB 16;  Length 297;
  Best Local Similarity   42.7%;  Pred. No. 2.3e-34;
  Matches  112;  Conservative   51;  Mismatches   73;  Indels   26;  Gaps    7;

Qy      84 RAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
           : ||||       ||  : :|:|:|  : |||  |: || |: ::| : || | | |||:
Db      49 KLADAASAETLPRFRTGIAVINKQDGGYDPVTEGDQAAETAIRALIEERFPQHGILGEEH 108

Qy     142 GWRCAEN---SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
           |     |     | :|||:||||||::||:| |::||||    :|:  :|::|||  :||:
Db     109 G-----NVGLDRDHIWVIDPIDGTRAFISGVPVWGTLIGFQSSGRATMGIMDQPFTKERY 163

Qy     199 IGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
           |||     |    ::|  | |  |: | |:||:||:| |: :  : :|:::|::  |
Db     164 F-ADGKAAWYFGPDGEKKIRTRDCASLSDAVLFTTTPHIFTAEEKPLYEKVQDQVRLFRY 222
```

```
Qy     255 GCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRP 314
           | ||||| |||:| ||:|:||||||||  :|:|||| |||::| | |           ||
Db     223 GVDCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQ^GGTMTTWDG----------GRP 272

Qy     315 -TSFNVVAAGDARVHKEALDAL 335
             ::|||   ||:|||  |
Db     273 ENGGRILAAGSKAVHEEALAIL 294


RESULT 4
Q930E4
ID   Q930E4      PRELIMINARY;      PRT;   308 AA.
AC   Q930E4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative inositol monophosphatase.
GN   RA0254 OR SMA0483.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OG   Plasmid pSymA (megaplasmid 1).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396509; PubMed=11481432;
RA   Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA   Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA   Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA   Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA   Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT   "Nucleotide sequence and predicted functions of the entire
RT   Sinorhizobium meliloti pSymA megaplasmid.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR   EMBL; AE007219; AAK64912.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
KW   Plasmid; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   308 AA;  32657 MW;  C8FE70A1785F9E96 CRC64;

  Query Match            27.9%;  Score 483;  DB 16;  Length 308;
  Best Local Similarity  39.6%;  Pred. No. 6e-31;
  Matches  107;  Conservative   38;  Mismatches  103;  Indels   22;  Gaps    4;

Qy      70 VGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILK 129
           :|| |:      |   || | : :      | | | | | |||  ||  |: :   |
Db      56 LGEEFAS-----FAHDIADLARQTISSAAGVRREPIAKSDASPVTETDRAVEKCLRERIAD 110

Qy     130 SFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
            || | : ||| |     |  :||||:|||||||:|: | |::||||:|   | |::|:|
Db     111 HFPDHGVLGEEFGAEGLGN--EFVWVIDPIDGTKAFVAGLPVYGTLISLTRGGTPILGLI 168

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           | |:  :||:|| |: ||||   |   |   || |::   :|  |    :     :|
Db     169 DNPMTGDRWLGVSGQPTTLNNVPIRTASTTALATAFIANGNPDAFSPADKSRVESLRRIT 228

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
           :  :||  | ||  :| | ||| :: || |||: :||||| |||| ||||:|
Db     229 RWCVYGGSCIAYGRVADGSVDISIDGGLDPYDYCALVPVITGAGGCITDWQG-------- 280

Qy     310 AESRPTSFN----VVAAGDARVHKEALDAL 335
           || : |     ||     :|:  |: |
Db     281 ---RPLTLNSGGLCVATATDLLHRHVLEIL 307


RESULT 5
Q9A643
ID   Q9A643      PRELIMINARY;      PRT;   278 AA.
AC   Q9A643;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Inositol monophosphatase family protein.
GN   CC2252.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005895; AAK24223.1; -.
DR   TIGR; CC2252; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
KW   Complete proteome.
SQ   SEQUENCE   278 AA;  29688 MW;  281299D0F028757A CRC64;

  Query Match            27.6%;  Score 477;  DB 16;  Length 278;
  Best Local Similarity  40.8%;  Pred. No. 1.6e-30;
  Matches  116;  Conservative   48;  Mismatches   88;  Indels   32;  Gaps    9;

Qy      73 LATERLVEVAQRAAD---AAGEVLRKYFRQRVEIIDK--------EDHS---PVTIADRE 118
           |: :||  :     |   |: :|:   ||    : ||        : |:   ||| |||
Db       3 LSADRLAALDAFILDLNRASADVILPLFRADHGLEDKGAGKNLPRDTHAAFDPVTEADRG 62

Qy     119 AEEAMVSVILKSFPTHAIFGEENGWRCAEN--SADFVWVLDPIDGTKSFITGKPLFGTLI 176
           || |: ::| : || | : ||| |    |:   |:||||||||||::|| | ||: |||
Db      63 AEAAIRALIAQRFPDHGVIGEEYG----EDRPDAEFVWVLDPIDGTRAFIAGLPLWTTLI 118

Qy     177 ALLHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQE---ISVRSCNLLAQAYLYTTSPH- 232
           | | |:||:| | || : | :||  |    ::| |   | || |  :  | : || |
Db     119 GLRHEGRPVLGSIGQPYVNEIFIGHAGGARLVSGGEARPIRVRECANINDAVIATTDPDA 178

Qy     233 LFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGA 292
           |:     |:::||   |:   ||| ||||::| | :|:|:|:|:||| :|  : :|:||||
Db     179 CFDGAERGAWLQVRAAAKLARLGCDAYAYAMVAMGKMDMVIEAGLKSWDIEAAIPLIEGA 238

Qy     293 GGSITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           || :|:|||:    ||     |    :| :|| |   ||| :||
Db     239 GGMVTNWRGE----PV--- GPNGGQMVISGDRRPLDEALVSLR 274


RESULT 6
Q8YDX6
ID   Q8YDX6      PRELIMINARY;      PRT;   272 AA.
AC   Q8YDX6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Extragenic suppressor protein SUHB.
GN   BMEII0048.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
```

```
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR   EMBL; AE009643; AAL53289.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   InterPro; IPR000146; In_FB_phphtase.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00377; INFBPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   272 AA;  29222 MW;  42C32EAF608C7E26 CRC64;

  Query Match           26.8%;  Score 463.5;  DB 16;  Length 272;
  Best Local S milarity  41.3%;  Pred. No. 1.9e-29;
  Matches  116;  Conservative   47;  Mismatches   93;  Indels   25;  Gaps    9;

Qy      66 RAAAVGELATERLVEVA---QRAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAE 120
           ||   |   |  |:: |   : || || : | : |||  |: :|      ||| ||| ||
Db       4 RAPCFGNRRTGLLIDKAFFSEVAAAAAAQTLPR-FRQLTEVDNKYSVGFDPVTEADRAAE 62

Qy     121 EAMVSVILKSFPTHAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALL 179
           |: :|| ::|| | | ||| |   |||:    ||::||:|||::||:| |::|||: |
Db      63 RAIRAVIGRTFPYHGILGEEYG---AENTDRSHVWIIDPVDGTRAFISGLPVWGTLVGLT 119

Qy     180 HNGKPVIGVIDQPILRERWIGVDGKQTTLN-----GQEISVRSCNLLAQAYLYTTSPHLF 234
           :|    |:: ||   | :   ||    |       : : ||   :|  | |:||:| ||
Db     120 VDGDARAGMMSQPFTGELFYS-DGDGAYLQRGDGAPRRLCVRKNAVLEDATLFTTTPALF 178

Qy     235 EADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGG 294
           : |   || |: : |::  || ||||:|:|| ||||||||:||: ||  :|:|:|:|| |||
Db     179 KGDDRKAFDRLESAVRLSRYGVDCYAFAMLAGGFVDIVVEAGLQTYDIAALIPIIEQAGG 238

Qy     295 SITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           :|   |        ||      ::|||    :|: ||| |
Db     239 VVTRRDGG------PAEQ---GGDIVAAATPALHQAALDLL 270


RESULT 7
Q9KWB5
ID   Q9KWB5      PRELIMINARY;      PRT;   263 AA.
AC   Q9KWB5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Riorf69 protein.
GN   RIORF69.
OS   Agrobacterium rhizogenes.
OG   Plasmid pRi1724.
     Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
     Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT   "Analysis of unique variable region of a plant root inducing plasmid,
RT   pRi1724, by the construction of its physical map and library.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT   "Genome structure of Ri plasmid (1):Construction of linking library
RT   and physical map of pRi1724 in Japanese Agrobacterium.";
RL   Nucleic Acids Symp. Ser. 39:189-190(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA   Yoshida K.;
RT   "The complete nucleotide sequence of a Ri (root inducing) plasmid
RT   indicates its chimerical structure between Ti and Sym plasmids.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT   "Analysis of unique variable region of a plant root inducing plasmid,
RT   pRi1724, by the construction of its physical map and library.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RX   MEDLINE=20241294; PubMed=10780382;
RA   Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA   Yoshida K.;
RT   "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
RT   its flanking regions of pRi1724 in Japanese Agrobacterium
RT   rhizogenes.";
RL   Nucleic Acids Symp. Ser. 42:67-68(1999).
DR   EMBL; AB039932; BAA97780.1; -.
DR   EMBL; AP002086; BAB16188.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Plasmid.
SQ   SEQUENCE   263 AA;  28110 MW;  2979EA9F8958A25F CRC64;

  Query Match           26.6%;  Score 460;  DB 2;  Length 263;
  Best Local Similarity  38.8%;  Pred. No. 3.4e-29;
  Matches  107;  Conservative   55;  Mismatches   82;  Indels   32;  Gaps    8;

Qy      74 ATERLVE---VAQRAADAAGEVLRKYFRQRVEIIDK--EDHSPVTIADREAEEAMVSVIL 128
           :|| ::    : : || || | | : ||    :::|      ||| |||| |:|: ::|
Db       3 STEDVISPDFLRRLAAVAARETLPR-FRLPGTVVNKIAGGFDPVTEADRETEKALRALIR 61

Qy     129 KSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGV 188
           :| | : ||| |   |:  |: |||:||||||:|||:| ||:|||: |  :|  | |:
Db      62 SEYPGHGVLGEEYGNEGAD--AELVWVIDPIDGTRSFISGIPLWGTLVGLTRDGDAVAGM 119

Qy     189 IDQPILRERWIGVDGKQTTLNG---------QEISVRSCNLLAQAYLYTTSPHLFEADAE 239
           : || : | :        :| ||          :::  :   || | |:||:| :||||
Db     120 MAQPFIGELF------YSTGNGAFYEGPHGQTQLATSATKDLADATLFTTTPAIFEADRR 173

Qy     240 DAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDW 299
           ::  :  :|::  || ||||| :||:| :| |||:|| ||| ::|:|:|| ||| :|||
Db     174 SSYDLLEKQVRLARYGTDCYAYCMLAAGQIDCVVEAGLYPYDIVALIPIIEQAGGVVTDW 233

Qy     300 RGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           |        ||       :||| :| :| : :  |
Db     234 DGG------PAE---RGGGIVAAANATLHAQIMACL 260


RESULT 8
Q98H43
ID   Q98H43      PRELIMINARY;      PRT;   262 AA.
AC   Q98H43;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Mono-phosphatase.
GN   MLL3039.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
```

```
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003001; BAB50023.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   262 AA;  27882 MW;  E38C73FB36D27CE5 CRC64;

  Query Match           26.0%;  Score 449;  DB 16;  Length 262;
  Best Local Similarity  39.8%;  Pred. No. 2.6e-28;
  Matches  105;  Conservative   50;  Mismatches   91;  Indels   18;  Gaps    7;

Qy      79 VEVAQRAADAAGEVLRKYFRQRVEIIDKEDHS--PVTIADREAEEAMVSVILKSFPTHAI 136
           ::  :| | ||       || :  : :||  |  ||| |||| | |: ::|   :| | |
Db      10 IDFMRRIAQAAAAETLPRFRAQGAVANKEKGSFDPVTEADRETERAIRALIAAQYPDHGI 69

Qy     137 FGEENGWRCAEN-SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |||:|   :|| |:  |||:||||||::||:| |::|||: |  :|  | |:: ||
b       70 LGEEHG---SENISSRHVWVIDPIDGTRAFISGLPVWGTLVGLTVDGDAVAGMMAQPFTG 126

Qy     196 ER-WIGVDGKQTTLNG--QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVP 252
           |  :    |      |  :::| |    | :| |:||:| ||: :|   :     :|::
Db     127 ELFYANASGSHYEGPGGPRKLSTRKTTRLDEATLFTTTPALFKGEARTRYDAFEKQVQLA 186

Qy     253 LYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAES 312
           || ||||:|::||| |||| : |||||| ::|:|:|| ||| :| : |          ||
Db     187 RYGADCYAFAMIASGSVDIVADPGLKPYDIVALIPIIEKAGGVVTTFDGG------PAEK 240

Qy     313 RPTSFNVVAAGDARVHKEALDALR 336
               :|:||    :|  |: |||
Db     241 ---GGDVLAAATPELHAAAMAALR 261


RESULT 9
P73806
ID   P73806      PRELIMINARY;      PRT;   267 AA
AC   P73806;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Extragenic suppressor.
GN   SUHB OR SLL1959.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90909; BAA17860.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   267 AA;  29243 MW;  3E0A3536C64AE707 CRC64;

  Query Match           22.9%;  Score 396.5;  DB 16;  Length 267;
  Best Local Similarity  36.9%;  Pred. No. 4.7e-24;
  Matches  101;  Conservative   49;  Mismatches   99;  Indels   25;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR----QRVEIIDKEDHSPVTIADREAEEAMVSVILKSF 131
           |: : :||: |  :||:| :|||    |    ||:   : || || |||:||| :|   |
Db       6 EQRLFIAQQLAAVSGEILIQYFRRSHLQGGTKIDQVS-AIVTQADEEAEQAMVDLIQAQF 64

Qy     132 PTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH-NGKPVIGVID 190
           |   :  || |   |  |  : |||||||| ||: | |:| ||| |:  : :||:|:
Db      65 PQDGVIREE-GKNIAGKSG-YTWVLDPIDGTSSFVRGLPIFATLIGLVDADMRPVLGIAH 122

Qy     191 QPILRERWIGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADA-EDAFIRV 245
           |||  :|| || |:|: :||            || | | : :|:| :|    :     :
Db     123 QPISGDRWQGVQGEQSNVNGIPLVNPYKASEINLTA-ACIVSTTPLMFTTPVQQQKMADI 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVD---IVVESGLKPYDFLSLVPVIEGAGGSITDWRGD 302
             : :   :| ||: |   |||:     ::||: |  ||| :|:|:: ||    |||:|
Db     182 YRQCQRTAFGGDCFNYLSAASGWTAMPLVIVEADLNFYDFCALIPILTGANYCFTDWQGK 241

Qy     303 KLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           :|         | |  |||: : ::| | |  |:
Db     242 EL--------TPESTEVVASPNPKLHSEILAFLQ 267


RESULT 10
Q988Q3
ID   Q988Q3      PRELIMINARY;      PRT;   239 AA.
AC   Q988Q3;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   N-amidino-scyllo-inosamine-4-phosphate phosphatase.
GN   MLL6642.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003009; BAB52894.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
KW   Complete proteome.
SQ   SEQUENCE   239 AA;  26168 MW;  EBFA2525C455B763 CRC64;

  Query Match           18.4%;  Score 317;  DB 16;  Length 239;
  Best Local Similarity  37.4%;  Pred. No. 1.1e-17;
  Matches   72;  Conservative   35;  Mismatches   72;  Indels   14;  Gaps    5;

Qy      84 RAADAAGEVLRKYFRQRVEIIDKED----HSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           | ||||  |   |:|: :|: :|       ||||||::||: : ::|   || ||: ||
Db      12 RLADAAANVALAYYRKPIEVEEKVKPGYRFDPVTIADKKAEQVIRALISGEFPDHAVLGE 71

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRER-W 198
           | |          |::||||||: || | |::|||| |  :||  :|:: ||   || |
Db      72 EFG---ESGQGPVKWIIDPIDGTRPFICGIPVWGTLIGLTVDGKACMGMMSQPFTGERFW 128

Qy     199 IGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
           | :   |      :  ::   || | |:| ||  |    :  | |:::   :  |
Db     129 --ATGTEAWTGGASGSSRLQTKAAGSLATAILHTNSPDRFPDFPDINFARLKDAALMTRY 186

Qy     255 GCDCYAYALLASG 267
           | :|||:|:||:|
Db     187 GGECYAFAMLAAG 199


RESULT 11
Q53743
ID   Q53743      PRELIMINARY;      PRT;   273 AA.
```

```
AC   Q53743;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Mono-phosphatase.
GN   PUR3.
OS   Streptomyces lipmanii (Streptomyces alboniger).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC12461;
RX   MEDLINE=96139493; PubMed=8576156;
RA   Tercero J.A., Espinosa J.C., Lacalle R.A., Jimenez A.;
RT   "The biosynthetic pathway of the aminonucleoside antibiotic Puromycin,
RT   as deduced from the molecular analysis of the pur cluster of
RT   Streptomyces alboniger.";
RL   J. Biol. Chem. 271:1579-1590(1996).
DR   EMBL; X92429; CAA63164.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
SQ   SEQUENCE   273 AA;  29090 MW;  C68A5E6240DF38DB CRC64;

  Query Match           18.3%;  Score 317;  DB 2;  Length 273;
  Best Local Similarity 32.3%;  Pred. No. 1.3e-17;
  Matches   86;  Conservative   43;  Mismatches  115;  Indels   22;  Gaps    6;

Qy      79 VEVAQRAADAAGEVLRKYFRQR-VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIF 137
           :|:| | || |  :  : |: |      | | :||| ||   | |:   :  : |
Db       7 LELAMRLADTADRITTRRFQARDCGYARKPDRTPVTDADTTVEAAVRESVRAARPDDDFA 66

Qy     138 GEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRER 197
           ||| |    | :|   |::|||||||:|: | |:: |||||| :|:| :||:  | || |
Db      67 GEETG---GEVTAGRTWIVDPIDGTKNFLRGVPVWATLIALLEDGRPTVGVVAAPALRSR 123

Qy     198 WIGVDGKQTTLN----GQE---ISVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKV 249
           |    |    |     | |   : |     |  ||| ||:   ::|  :  |::|: :
Db     124 WWAAAGHGAWLRRGSAGAEPLRLHVSGVARLENAYLSTTNTRTWDAFHSRAAYLRLADAC 183

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
               |   : ::| | |||  |  : |:|  :|  ::| |||  ||  |
Db     184 WEDRAFGDFLQHCMVAEGTVDIAAEPVVSPWDIAALQILVEEAGGVCTDLLG-------- 235

Qy     310 AESRPTSFNVVAAGDARVHKEALDAL 335
               |      : :  :|: |::||
Db     236 --GSPQRGTGALSANPELHRLAVEAL 259


RESULT 12
Q8RJW8
ID   Q8RJW8      PRELIMINARY;      PRT;   268 AA.
AC   Q8RJW8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative monophosphatase.
GN   ATAP3.
OS   Streptomyces capreolus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Actinosynnemataceae;
OC   Saccharothrix.
OX   NCBI_TaxID=66854;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL3817;
RX   MEDLINE=95255257; PubMed=7737149;
RA   Barrasa M.I., Tercero J.A., Lacalle R.A., Jimenez A.;
RT   "The ard1 gene from Streptomyces capreolus encodes a polypeptide of
RT   the ABC-transporters superfamily which confers resistance to the
RT   aminonucleoside antibiotic A201A.";
RL   Eur. J. Biochem. 228:562-569(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL3817;
RA   Jimenez A.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X84374; CAD27649.1; -.
SQ   SEQUENCE   268 AA;  28723 MW;  52586EF8AA22778E CRC64;

  Query Match           17.5%;  Score 302.5;  DB 2;  Length 268;
  Best Local Similarity 31.5%;  Pred. No. 1.9e-16;
  Matches   87;  Conservative   42;  Mismatches   98;  Indels   49;  Gaps   10;

Qy      81 VAQRAADAAGEVLRKYFRQRVEIID-KEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   || :  | |  || |   :| | | :||| ||   |:| |    :: |   | ||
Db      19 LALELADLSDAVTRSRFRARDLRVDRKPDRTPVTDADLAVEQA-VRTAWRARPRDRIAGE 77

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWI 199
           | |    :      ||:||||||:|: | |:: | :||: :|:| :||:  | || ||
Db      78 EGGGAITDRR---TWVVDPIDGTKNFLRGVPVWATSLALVEDGRPTVGVVSAPALRRRWW 134

Qy     200 GVDG----KQTTLNGQEISVRSCNLLAQAYLYTTSP---HLFEADAEDAFIRVRNKVKVP 252
           |    :      : |   : |  ||| |||     || : |  |::|: :
Db     135 AAAGHGAWAREDRGTARLHVSGVDRLEAAYLSTTSLSTWDLFHSRA--AYLRLAD----- 187

Qy     253 LYGC-------DCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
           |       |   : |:| | :||  |  : |:|  ::  ::| |||  |  ||:
Db     188 --ACWEDRGFGDFLQHCLVAEGLIDIAAEPVVSPWDVAAVQVIVEEAGGRCTGLRGE--- 242

Qy     306 WPVTAESRPTSFNVVAAG------DARVHKEALDAL 335
                  ||: ||      :  :| :|:  |
Db     243 -----------NVLGAGGGALSANPHLHGQAVALL 266


RESULT 13
Q8ZCR8
ID   Q8ZCR8      PRELIMINARY;      PRT;   267 AA.
AC   Q8ZCR8;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative inositol monophosphatase family protein.
GN   YPO2899.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414154; CAC92150.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   InterPro; IPR000146; In_FB_phphtase.
DR   Pfam; PF00459; inositol_P; 1.
DR   PRINTS; PR00377; INFBPHPHTASE.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   267 AA;  29155 MW;  B2AD807A0778A78C CRC64;

  Query Match           16.9%;  Score 293;  DB 16;  Length 267;
  Best Local Similarity 31.6%;  Pred. No. 1.1e-15;
  Matches   86;  Conservative   43;  Mismatches  111;  Indels   32;  Gaps    8;
```

```
Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || :: | :    ||   |  :  || ||:||  :: || ||:| |
Db       4 MLTIAIRAARKAGNLIAKNYETPDAVEASQKGSNDFVTNVDRDAEHLIIDVIRKSYPKHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |    |:  |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db      64 IISEECGELVGEDD-DVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCN-------LLAQAYLYTTSPHLFEADAEDAFIRVRNK 248
           | :    |:   |||  :  |  |       :||  : :    |       |:|||  |
Db     123 ELFTASRGQGAQLNGYRL--RGTNAKDLDGTILATGFPFKVKQH------APAYIRVVGK 174

Qy     249 V-----KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           :          |      | :|:| ||   | ||||:||     ::  :|| :||: |
Db     175 LFEQCADFRRTGSAALDLAYVAAGRVDGFFEIGLKPWDFAGGELLVRESGGIVTDFAGGH 234

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
            |:       :| |:| ||: |: |  : |:
Db     235 NHF--------SSGNIV-AGNPRIVKSIVQAM 257
```

RESULT 14
O32889
ID O32889 PRELIMINARY; PRT; 255 AA.
AC O32889;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MONOPHOPHATASE (Putative monophosphatase).
GN ML0662 OR MLCB1779.29.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; Z98271; CAB11010.1; -.
DR EMBL; AL583919; CAC30171.1; -.
DR Leproma; ML0662; -.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
KW Complete proteome.
SQ SEQUENCE 255 AA; 27040 MW; 2633DDC72324D545 CRC64;

```
  Query Match            16.4%;  Score 283;  DB 16;  Length 255;
  Best Local Similarity  31.4%;  Pred. No. 6.5e-15;
  Matches   85;  Conservative   40;  Mismatches  102;  Indels   44;  Gaps    9;

Qy      81 VAQRAADAAGEVLRKYF-RQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   || | :    |    : :  | | :||| |||  |  : :|: :  |   | ||
Db       2 LALTLADRADALTSARFGALNLRVDTKPDLTPVTDADRAVEADVRAVLGRERPKDGILGE 61

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW- 198
           | |     :     |::|||||||:|: | |:: :|||||:| | |||:  | |: ||
Db      62 EYGGTITFSGQQ--WIVDPIDGTKNFVRGVPVWASLIALLEDGVPSIGVVSAPALQRRWW 119

Qy     199 --------IGVDG--------KQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAF 242
                  : |||        :   ||   :|  | :  ||  |             | |
```

```
Db     120 AARGQGAFVAVDGVPRRLAVSEVADLNSASLSFSSLSGWAQRGL-----------RDRF 167

Qy     243 IRVRNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
           : : : | :|  || |  :| ||| | :|:  |  :  :|  :|  |:  ||| :|   |
Db     168 LELTDAVWRVRAYG-DFLSYCLLAEGAIDVAAEPKVSVWDLAALDIVVREAGGVLTGLDG 226

Qy     302 DKLHWPVTAESRPTSFNVVAAGDARVHKEAL 332
                     |   : ||  : |:|:| |
Db     227 ---------TPGPHGGSAVAT-NGRLHQEVL 247
```

RESULT 15
Q9K4B1
ID Q9K4B1 PRELIMINARY; PRT; 266 AA.
AC Q9K4B1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative monophosphatase.
GN SCO5208 OR SC7E4.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL359214; CAB94593.1; -.
DR HSSP; P29218; 1IMF.
DR InterPro; IPR002106; AAtRNA_ligaseII.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS00629; IMP_1; 1.
SQ SEQUENCE 266 AA; 29083 MW; 9C2802DB177B243E CRC64;

```
  Query Match            16.1%;  Score 278.5;  DB 16;  Length 266;
  Best Local Similarity  30.3%;  Pred. No. 1.6e-14;
  Matches   83;  Conservative   51;  Mismatches  111;  Indels   29;  Gaps   10;

Qy      77 RLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAI 136
           ||  |   |||||   : ::    :::  | | :||: ||: ||| :   : :: |  ::
```

```
Db       9 RLAHVLADAADAA--TMDRFKALDLKVETKPDMTPVSEADKAAEELIRGHLSRARPRDSV 66

Qy     137 FGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH----NGKPVIGVIDQP 192
           ||| |      :    ||:|||||||::: | |:: ||||||:       :||:|::  |
Db      67 HGEEFG---VAGTGPRRWVIDPIDGTKNYVRGVPVWATLIALMEAKEGGYQPVVGLVSAP 123

Qy     193 ILRERWIGVD------GKQTTLNGQEISVRSCNLLAQA-YLYTTSPHLFEADAEDAFIRV 245
           |  ||  |:      |:  | :   : |   : |: | : |::     |    | |: :
Db     124 ALGRRWWAVEDHGAFTGRSLT-SAHRLHVSQVSTLSDASFAYSSLSGWEEQGRLDGFLDL 182

Qy     246 RNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKL 304
             :| :   || | : | ::| | ||:  |  |  :|  :   ::  |||: |   |
Db     183 TREVWRTRAYG-DFWPYMMVAEGSVDLCAEPELSLWDMAANAIIVTEAGGTFTGLDG--- 238

Qy     305 HWPVTAESRPTSFNVVAAGDARVHKEALDALRWR 338
                 | |  || : |:| | |  |  |
Db     239 ------RPGPHSGN-AAASNGRLHDELLGYLNQR 265
```

Search completed: November 17, 2002, 16:41:29
Job time : 95 secs

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2002, 17:13:04 ; Search time 28.74 Seconds
(without alignments)
2034.526 Million cell updates/sec

Title: US-09-686-522A-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1020	59.0	286	10	Q9T021	Q9t021 arabidopsis
2	551	31.8	259	16	Q92M31	Q92m31 rhizobium m
3	483	27.9	308	16	Q930E4	Q930e4 rhizobium m
4	477	27.6	278	16	Q9A643	Q9a643 caulobacter
5	460	26.6	263	2	Q9KWB5	Q9kwb5 agrobacteri
6	449	26.0	262	16	Q98H43	Q98h43 rhizobium l
7	396.5	22.9	267	16	P73806	P73806 synechocyst
8	317	18.3	239	16	Q988Q3	Q988q3 rhizobium l
9	317	18.3	273	2	Q53743	Q53743 streptomyce
10	292	16.9	271	16	Q9HXI4	Q9hxi4 pseudomonas
11	287	16.6	267	16	Q9CNV8	Q9cnv8 pasteurella
12	283	16.4	255	16	O32889	O32889 mycobacteri
13	279	16.1	264	16	O67791	O67791 aquifex aeo
14	278.5	16.1	266	2	Q9K4B1	Q9k4b1 streptomyce
15	273.5	15.8	260	16	P95189	P95189 mycobacteri
16	265.5	15.3	271	10	Q9M8S8	Q9m8s8 arabidopsis
17	258.5	14.9	371	10	Q94F00	Q94f00 arabidopsis
18	248	14.3	288	16	Q9KTY5	Q9kty5 vibrio chol
19	239.5	13.8	266	17	Q9HQ12	Q9hq12 halobacteri
20	237	13.7	272	16	Q92RQ1	Q92rq1 rhizobium m
21	234	13.5	275	16	Q9PAM0	Q9pam0 xylella fas
22	231.5	13.4	353	10	Q9SA15	Q9sa15 arabidopsis
23	230	13.3	221	2	P71487	P71487 methylomicr
24	230	13.3	287	16	P74542	P74542 synechocyst
25	229	13.2	264	16	Q98D39	Q98d39 rhizobium l
26	228	13.2	266	16	Q98F59	Q98f59 rhizobium l
27	221	12.8	266	16	Q92WJ1	Q92wj1 rhizobium m
28	221	12.8	272	2	O30546	O30546 agrobacteri
29	220	12.7	275	16	Q98NP3	Q98np3 rhizobium l
30	220	12.7	294	16	Q92X03	Q92x03 rhizobium m
31	219	12.7	290	11	Q91UZ5	Q91uz5 mus musculu
32	218	12.6	319	5	Q23493	Q23493 caenorhabdi
33	215	12.4	255	2	Q9R6D4	Q9r6d4 agrobacteri
34	213	12.3	264	2	Q9S3X5	Q9s3x5 streptomyce
35	212.5	12.3	277	11	Q9D066	Q9d066 mus musculu
36	212	12.3	284	5	Q9VUW2	Q9vuw2 drosophila
37	211	12.2	266	16	Q92M71	Q92m71 rhizobium m
38	210.5	12.2	288	5	Q9VUW4	Q9vuw4 drosophila
39	209	12.1	256	2	P72460	P72460 streptomyce
40	209	12.1	257	16	Q92CW7	Q92cw7 listeria in
41	208.5	12.1	277	11	Q924B0	Q924b0 mus musculu
42	207.5	12.0	127	2	O69216	O69216 azotobacter
43	207	12.0	254	17	Q9V1Z6	Q9v1z6 pyrococcus
44	207	12.0	397	10	Q9M0Y6	Q9m0y6 arabidopsis
45	206.5	11.9	287	2	Q52855	Q52855 rhizobium l

ALIGNMENTS

RESULT 1
Q9T021
ID Q9T021 PRELIMINARY; PRT; 286 AA.
AC Q9T021;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel 13, Last sequence updat:)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation upd:te)
DE INOSITOL MONOPHOSPHATASE-LIKE PROTEIN.
GN T22F8.20 OR AT4G39120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Grueneisen A., Wambutt R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL050351; CAB43627.1; -.
DR EMBL; AL161594; CAB80575.1; -.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
SQ SEQUENCE 286 AA; 31635 MW; 5C2BB2BA90FD10BB CRC64;

```
  Query Match             59.0%;  Score 1020;  DB 10;  Length 286;
  Best Local Similarity   69.5%;  Pred. No. 9.2e-75;
  Matches  182;  Conservative   45;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :|   |    |||:|||:|||||:: :|:||:| |||||||: |||||||:| :: |:||
Db      24 DRFAAVGNALADASGEVIRKYFRKKFDIVDKDDMSPVTIADQMAEEAMVSIIFQNLPSHA 83

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |:||| |||| | |||:||||||||||||||||:|||||||: |||::|:||||||:
Db      84 IYGEEKGWRCKEESADYVWVLDPIDGTKSFITGKPVFGTLIALLYKGKPILGLIDQPILK 143

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           |||||::|::| |||::|| |||  |:||||||||||||  :|| |: |||:||||||||
Db     144 ERWIGMNGRRTKLNGEDISTRSCPKLSQAYLYTTSPHLFSEEAEKAYSRVRDKVKVPLYG 203

Qy     256 CDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPT 315
           |||||||||||||||:|:||||||||||:|||||||||:|||| | :  |  :: :  |
Db     204 CDCYAYALLASGFVDLVIESGLKPYDFLALVPVIEGAGGTITDWTGKRFLWEASSSAVAT 263

Qy     316 SFNVVAAGDARVHKEALDALRW 337
           |||||||||: :|::||::| |
Db     264 SFNVVAAGDSDIHQQALESLEW 285


RESULT   2
Q92M31
ID   Q92M31      PRELIMINARY;      PRT;   259 AA.
AC   Q92M31;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE MONOPHOSPHATASE PROTEIN.
GN   SMC04042.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21368234; PubMed=11474104;
RA   Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA   Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA   Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA   Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA   Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
     Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
     Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA   Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA   Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
RA   Vorhoelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT   "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL   Science 293:668-672(2001).
DR   EMBL; AL591791; CAC47404.1; -.
KW   Complete proteome.
SQ   SEQUENCE   259 AA;  28162 MW;  2062305CB4B3C871 CRC64;


  Query Match             31.8%;  Score 551;  DB 16;  Length 259;
  Best Local Similarity   44.0%;  Pred. No. 8.7e-37;
  Matches  117;  Conservative   48;  Mismatches   73;  Indels   28;  Gaps    7;

Qy      84 RAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
           | ||||       ||    :::| :    ||| ||| || :: ::|  ||| | | |||:
Db      10 RLADAAKAETMPRFRVGTSVLNKLEGGFDPVTEADRSAEASIRALIESSFPEHGILGEEH 69

Qy     142 GWRCAEN---SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
           |     |     : |||:||||||::||:| |::|||| |  ||| |:|::|||    ||:
Db      70 G-----NIGLDRELVWVIDPIDGTRAFISGLPVWGTLIGLYRNGKAVMGLMDQPFTGERF 124

Qy     199 IGVDGKQTTLNGQE----ISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
```

```
           ||::    | :    :: | |: |: | |:||||||:  : :: |  :: ||::  |
Db     125 F-ADGEKALYRGPDGERVLATRPCHALSDAVLFTTSPHLYTGELKERFEALQEKVRLFRY 183

Qy     255 GCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRP 314
           ||||||:||||:| ||:||| ||||||   |:|:|| ||| ||||:|           |
Db     184 GCDCYAFALLAAGHVDLVVECGLKPYDVGGLIPLIEQAGGIITDWQGG----------P 232

Qy     315 TSF--NVVAAGDARVHKEALDALRWR 338
                 ::|||   :| :||:||: |
Db     233 AEMGGEIIAAGSRELHAQALEALKGR 258


RESULT   3
Q930E4
ID   Q930E4      PRELIMINARY;      PRT;   308 AA.
AC   Q930E4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE INOSITOL MONOPHOSPHATASE.
GN   SMA0483.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OG   Plasmid pSymA (megaplasmid 1).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396509; PubMed=11481432;
RA   Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA   Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA   Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA   Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA   Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT   "Nucleotide sequence and predicted functions of the entire
RT   Sinorhizobium meliloti pSymA megaplasmid.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR   EMBL; AE007219; AAK64912.1; -.
KW   Plasmid; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   308 AA;  32657 MW;  C8FE70A1785F9E96 CRC64;


  Query Match             27.9%;  Score 483;  DB 16;  Length 308;
  Best Local Similarity   39.6%;  Pred. No. 3.6e-31;
  Matches  107;  Conservative   38;  Mismatches  103;  Indels   22;  Gaps    4;

Qy      70 VGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILK 129
           :|| |:       |   || | : :       | | | | | ||||  ||  |: :   |
Db      56 LGEFAS-----FAHDIADLARQTISSAAGVRREPIAKSDASPVTETDRAVEKCLRERIAD 110

Qy     130 SFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           || | : ||| |      |  :||||:|||||||:|: | |::||||:|   | |::|:|
Db     111 HFPDHGVLGEEFGAEGLGN--EFVWVIDPIDGTKAFVAGLPVYGTLISLTRGGTPILGLI 168

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           | |:  :||:|| |: ||||   |   |   || |::   :|  |    :     :|
Db     169 DNPMTGDRWLGVSGQPTTLNNVPIRTASTTALATAFIANGNPDAFSPADKSRVESLRRIT 228

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
           :  :||  | ||  :| | ||| :: || |||: :||||| |||| ||||:|
Db     229 RWCVYGGSCIAYGRVADGSVDISIDGGLDPYDYCALVPVITGAGGCITDWQG-------- 280

Qy     310 AESRPTSFN----VVAAGDARVHKEALDAL 335
           || : |     ||     :|:  |: |
Db     281 ---RPLTLNSGGLCVATATDLLHRHVLEIL 307


RESULT   4
Q9A643
ID   Q9A643      PRELIMINARY;      PRT;   278 AA.
AC   Q9A643;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   INOSITOL MONOPHOSPHATASE FAMILY PROTEIN.
GN   CC2252.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=69394;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005895; AAK24223.1; -.
DR   TIGR; CC2252; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 2.
KW   Complete proteome.
SQ   SEQUENCE   278 AA;  29688 MW;  281299D0F028757A CRC64;
```

```
  Query Match           27.6%;  Score 477;  DB 16;  Length 278;
  Best Local Similarity  40.8%;  Pred. No. 9.6e-31;
  Matches  116;  Conservative   48;  Mismatches   88;  Indels   32;  Gaps    9;

Qy      73 LATERLVEVAQRAAD---AAGEVLRKYFRQRVEIIDK--------EDHS---PVTIADRE 118
           |: :|| :     |   |: :|:   ||     : ||         : |:   ||| |||
Db       3 LSADRLAALDAFILDLNRASADVILPLFRADHGLEDKGAGKNLPRDTHAAFDPVTEADRG 62

Qy     119 AEEAMVSVILKSFPTHAIFGEENGWRCAEN--SADFVWVLDPIDGTKSFITGKPLFGTLI 176
           || |: ::| : || | : ||| |    |:   |:||||||||||::|| | ||: |||
Db      63 AEAAIRALIAQRFPDHGVIGEEYG----EDRPDAEFVWVLDPIDGTRAFIAGLPLWTTLI 118

Qy     177 ALLHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQE---ISVRSCNLLAQAYLYTTSPH- 232
            | | |:||:| | || : | :|| |    ::| |   | || |  :  | : || |
Db     119 GLRHEGRPVLGSIGQPYVNEIFIGHAGGARLVSGGEARPIRVRECANINDAVIATTDPDA 178

Qy     233 LFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGA 292
           |:     |:::||   |:   ||| ||||::| | :|:|:|:||| :|  : :|:||||
Db     179 CFDGAERGAWLQVRAAAKLARLGCDAYAYAMVAMGKMDMVIEAGLKSWDIEAAIPLIEGA 238

Qy     293 GGSITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           || :|:|||:    ||     |    :| :|| |   ||| :||
Db     239 GGMVTNWRGE----PV----GPNGGQMVISGDRRPLDEALVSLR 274
```

```
RESULT   5
Q9KWB5
ID   Q9KWB5      PRELIMINARY;      PRT;   263 AA.
AC   Q9KWB5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RIORF69 PROTEIN.
GN   RIORF69.
OS   Agrobacterium rhizogenes.
OG   Plasmid pRi1724.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT   "Analysis of unique variable region of a plant root inducing plasmid,
RT   pRi1724, by the construction of its physical map and library.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT   "Genome structure of Ri plasmid (1):Construction of linking library
RT   and physical map of pRi1724 in Japanese Agrobacterium.";
RL   Nucleic Acids Symp. Ser. 39:189-190(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA   Yoshida K.;
RT   "The complete nucleotide sequence of a Ri (root inducing) plasmid
RT   indicates its chimerical structure between Ti and Sym plasmids.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RA   Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT   "Analysis of unique variable region of a plant root inducing plasmid,
RT   pRi1724, by the construction of its physical map and library.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724; PLASMID=PRI1724;
RX   MEDLINE=20241294; PubMed=10780382;
RA   Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA   Yoshida K.;
RT   "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
RT   its flanking regions of pRi1724 in Japanese Agrobacterium
RT   rhizogenes.";
RL   Nucleic Acids Symp. Ser. 42:67-68(1999).
DR   EMBL; AB039932; BAA97780.1; -.
DR   EMBL; AP002086; BAB16188.1; -.
DR   HSSP; P29218; 11MF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Plasmid.
SQ   SEQUENCE   263 AA;  28110 MW;  2979EA9F8958A25F CRC64;
```

```
  Query Match           26.6%;  Score 460;  DB 2;  Length 263;
  Best Local Similarity  38.8%;  Pred. No. 2.1e-29;
  Matches  107;  Conservative   55;  Mismatches   82;  Indels   32;  Gaps    8;

Qy      74 ATERLVE---VAQRAADAAGEVLRKYFRQRVEIIDK--EDHSPVTIADREAEEAMVSVIL 128
           :|| ::    : : || || | | : ||    :::|      ||| |||| |:|: ::|
Db       3 STEDVISPDFLRRLAAVAARETLPR-FRLPGTVVNKIAGGFDPVTEADRETEKALRALIR 61

Qy     129 KSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGV 188
             :| | : ||| |   |:  |: |||:||||||:|||:| ||:|||: |  :|  | |:
Db      62 SEYPGHGVLGEEYGNEGAD--AELVWVIDPIDGTRSFISGIPLWGTLVGLTRDGDAVAGM 119

Qy     189 IDQPILRERWIGVDGKQTTLNG--------QEISVRSCNLLAQAYLYTTSPHLFEADAE 239
           : || : | :       :| ||         :::  :   || | |:||:| :||||
Db     120 MAQPFIGELF------YSTGNGAFYEGPHGQTQLATSATKDLADATLFTTTPAIFEADRR 173

Qy     240 DAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDW 299
            ::  :  :|::  || ||||| :||:| :| |||:|| ||| ::|:|:|| ||| :|||
Db     174 SSYDLLEKQVRLARYGTDCYAYCMLAAGQIDCVVEAGLYPYDIVALIPIIEQAGGVVTDW 233

Qy     300 RGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           |       ||       :||| :| :| : :  |
Db     234 DGG------PAE---RGGGIVAAANATLHAQIMACL 260
```

```
RESULT   6
Q98H43
ID   Q98H43      PRELIMINARY;      PRT;   262 AA.
AC   Q98H43;
```

```
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   MONO-PHOSPHATASE.
GN   MLL3039.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
     DNA Res. 7:331-338(2000).
     EMBL; AP003001; BAB50023.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
KW   Complete proteome.
SQ   SEQUENCE   262 AA;  27882 MW;  E38C73FB36D27CE5 CRC64;
```

```
  Query Match             26.0%;  Score 449;  DB 16;  Length 262;
  Best Local Similarity   39.8%;  Pred. No. 1.7e-28;
  Matches  105;  Conservative   50;  Mismatches   91;  Indels   18;  Gaps    7;

Qy      79 VEVAQRAADAAGEVLRKYFRQRVEIIDKEDHS--PVTIADREAEEAMVSVILKSFPTHAI 136
           ::  :| | ||       || :  : :||  |  ||| |||| | |: ::|   :| | |
Db      10 IDFMRRIAQAAAAETLPRFRAQGAVANKEKGSFDPVTEADRETERAIRALIAAQYPDHGI 69

Qy     137 FGEENGWRCAEN-SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |||:|   :|| |:  |||:||||||::||:| |::|||: |  :|  | |:: ||
Db      70 LGEEHG---SENISSRHVWVIDPIDGTRAFISGLPVWGT AVGLTVDGDAVAGMMAQPFTG 126

Qy     196 ER-WIGVDGKQTTLNG--QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVP 252
           |  :   |      |  :::| |    | :| |:||:| ||: :|   :     :|::
Db     127 ELFYANASGSHYEGPGGPRKLSTRKTTRLDEATLFTTTPALFKGEARTRYDAFEKQVQLA 186

Qy     253 LYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAES 312
           || ||||:|::||| |||| : |||||| ::|:|:|| ||| :| : |        ||
Db     187 RYGADCYAFAMIASGSVDIVADPGLKPYDIVALIPIIEKAGGVVTTFDGG------PAEK 240

Qy     313 RPTSFNVVAAGDARVHKEALDALR 336
               :|:||    :|  |: |||
Db     241 ---GGDVLAAATPELHAAAMAALR 261
```

```
RESULT   7
P73806
ID   P73806      PRELIMINARY;      PRT;   267 AA.
AC   P73806;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   EXTRAGENIC SUPPRESSOR.
GN   SUHB OR SLL1959.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
```

```
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90909; BAA17860.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   267 AA;  29243 MW;  3E0A3536C64AE707 CRC64;
```

```
  Query Match             22.9%;  Score 396.5;  DB 16;  Length 267;
  Best Local Similarity   36.9%;  Pred. No. 3.1e-24;
  Matches  101;  Conservative   49;  Mismatches   99;  Indels   25;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR----QRVEIIDKEDHSPVTIADREAEEAMVSVILKSF 131
           |: : :||: |  :||:| :|||    |    ||:   : || || |||:||| :|   |
Db       6 EQRLFIAQQLAAVSGEILIQYFRRSHLQGGTKIDQVS-AIVTQADEEAEQAMVDLIQAQF 64

Qy     132 PTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH-NGKPVIGVID 190
           |   :  || |   | |  : |||||||| ||: | |:| ||| |:  : :||:|:
Db      65 PQDGVIREE-GKNIAGKSG-YTWVLDPIDGTSSFVRGLPIFATLIGLVDADMRPVLGIAH 122

Qy     191 QPILRERWIGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADA-EDAFIRV 245
           |||  :|| || |:|: :||            || | | : :|:| :|     :     :
Db     123 QPISGDRWQGVQGEQSNVNGIPLVNPYKASEINLTA-ACIVSTTPLMFTTPVQQQKMADI 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVD---IVVESGLKPYDFLSLVPVIEGAGGSITDWRGD 302
             : :   :| ||: |   |||:     ::||: |  ||| :|:|:: ||    |||:|
Db     182 YRQCQRTAFGGDCFNYLSAASGWTAMPLVIVEADLNFYDFCALIPILTGANYCFTDWQGK 241

Qy     303 KLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           :|         | |  |||: : ::| | |  |:
Db     242 EL--------TPESTEVVASPNPKLHSEILAFLQ 267
```

```
RESULT   8
Q988Q3
ID   Q988Q3      PRELIMINARY;      PRT;   239 AA.
AC   Q988Q3;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   N-AMIDINO-SCYLLO-INOSAMINE-4-PHOSPHATE PHOSPHATASE.
GN   MLL6642.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003009; BAB52894.1; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
KW   Complete proteome.
SQ   SEQUENCE   239 AA;  26168 MW;  EBFA2525C455B763 CRC64;
```

```
  Query Match             18.3%;  Score 317;  DB 16;  Length 239;
  Best Local Similarity   37.3%;  Pred. No. 7.3e-18;
  Matches   72;  Conservative   35;  Mismatches   72;  Indels   14;  Gaps    5;
```

```
Qy      84 RAADAAGEVLRKYFRQRVEIIDKED----HSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           | ||||  |   |:|: :|: :|        ||||||::||: : ::|   || ||: ||
Db      12 RLADAAANVALAYYRKPIEVEEKVKPGYRFDPVTIADKKAEQVIRALISGEFPDHAVLGE 71

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRER-W 198
           | |           |::|||||: || | |::|||| |  :||  :|:: ||   || |
Db      72 EFG---ESGQGPVKWIIDPIDGTRPFICGIPVWGTLIGLTVDGKACMGMMSQPFTGERFW 128

Qy     199 IGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
               | :    |      :  ::   || | |:| ||  |    :  | |:::   :  |
Db     129 --ATGTEAWTGGASGSSRLQTKAAGSLATAILHTNSPDRFPDFPDINFARLKDAALMTRY 186

Qy     255 GCDCYAYALLASG 267
           | :|||:|:||:|
Db     187 GGECYAFAMLAAG 199


RESULT   9
Q53743
ID   Q53743      PRELIMINARY;      PRT;   273 AA.
AC   Q53743;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MONO-PHOSPHATASE.
GN   PUR3.
OS   Streptomyces lipmanii (Streptomyces alboniger).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC12461;
RX   MEDLINE=96139493; PubMed=8576156;
RA   Tercero J.A., Espinosa J.C., Lacalle R.A., Jimenez A.;
RT   "The biosynthetic pathway of the aminonucleoside antibiotic Puromycin,
RT   as deduced from the molecular analysis of the pur cluster of
RT   Streptomyces alboniger.";
RL   J. Biol. Chem. 271:1579-1590(1996).
DR   EMBL; X92429; CAA63164.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 2.
DR   PROSITE; PS00629; IMP_1; 1.
SQ   SEQUENCE   273 AA;  29090 MW;  C68A5E6240DF38DB CRC64;


  Query Match             18.3%;  Score 317;  DB 2;  Length 273;
  Best Local Similarity   32.3%;  Pred. No. 8.9e-18;
  Matches   86;  Conservative   43;  Mismatches  115;  Indels   22;  Gaps    6;

Qy      79 VEVAQRAADAAGEVLRKYFRQR-VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIF 137
           :|:| | || |  :  : |: |      | | :||| ||   | |:   :  : |
Db       7 LELAMRLADTADRITTRRFQARDCGYARKPDRTPVTDADTTVEAAVRESVRAARPDDDFA 66

Qy     138 GEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRER 197
           ||| |    | :|   |::|||||||:|: | |:: |||||| :|:| :||:  | || |
Db      67 GEETG---GEVTAGRTWIVDPIDGTKNFLRGVPVWATLIALLEDGRPTVGVVAAPALRSR 123

Qy     198 WIGVDGKQTTLN----GQE---ISVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKV 249
           |   |    |     | |   : |     |  ||| ||:   ::|  :  |::|: :
Db     124 WWAAAGHGAWLRRGSAGAEPLRLHVSGVARLENAYLSTTNTRTWDAFHSRAAYLRLADAC 183

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
               |   : ::| | |||  |  : |:|  :|  ::| |||  ||  |
Db     184 WEDRAFGDFLQHCMVAEGTVDIAAEPVVSPWDIAALQILVEEAGGVCTDLLG-------- 235

Qy     310 AESRPTSFNVVAAGDARVHKEALDAL 335
               |       : :  :|: |::||
Db     236 --GSPQRGTGALSANPELHRLAVEAL 259


RESULT  10
```

```
Q9HXI4
ID   Q9HXI4      PRELIMINARY;      PRT;   271 AA.
AC   Q9HXI4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   EXTRAGENIC SUPPRESSOR PROTEIN SUHB.
GN   PA3818.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004799; AAG07205.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; 1.
KW   Complete proteome.
SQ   SEQUENCE   271 AA;  29521 MW;  EE1D43CFFEB871EA CRC64;


  Query Match             16.9%;  Score 292;  DB 16;  Length 271;
  Best Local Similarity   29.7%;  Pred. No. 9.4e-16;
  Matches   82;  Conservative   54;  Mismatches  106;  Indels   34;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVETI---DKEDHSPVTIADREAEEAMVSVILKSFP 132
           :: :| ||| :|||::   ||  :|:::|   :|:    ||  || ||: :|: : |::|
Db       4 MLNIALRAARSAGELI---FRSIERLDVISVNEKDAKDYVTEVDRAAEQTIVAALRKAYP 60

Qy     133 THAIFGEENGW-RCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
           |||| ||| |:   :   ||::||:||:||| :|| | | |   ||  : |:    |:
Db      61 THAIMGEEGGFIEGSGEGADYLWVIDPLDGTTNFIHGVPHFAVSIACKYKGRLEHAVVLD 120

Qy     192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHL-FEADAEDAFIR-----V 245
           |: :| :    |:  |||: : |    | | | |   |  : | | :: |
Db     121 PVRQEEFTASRGRGAALNGFRLRVSGRKSLEGALLGTGFPFRDNQIDNILDNYLNMFRSLV 180

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG---- 301
               :   |      | :|:|  |   | ||  :|  :   ::: ||| ::|: |
Db     181 GQTAGIRRAGAASLDLAYVAAGRYDAFWEFGLSEWDMAAGALLVQEAGGLVSDFTGSHEF 240

Qy     302 -DKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
            :| |              : ||: :  |  |  ::
Db     241 LEKGH-------------IVAGNTKCFKALLTTIQ 262


RESULT  11
Q9CNV8
ID   Q9CNV8      PRELIMINARY;      PRT;   267 AA.
AC   Q9CNV8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SUHB.
GN   SUHB OR PM0315.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006067; AAK02399.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   267 AA;  29470 MW;  E7070FAE3BA589CA CRC64;


  Query Match             16.6%;  Score 287;  DB 16;  Length 267;
  Best Local Similarity   30.6%;  Pred. No. 2.3e-15;
  Matches   83;  Conservative   49;  Mismatches  111;  Indels   28;  Gaps    6;

Qy       78 LVEVAQRAADAAGEVLRKYFRQR--VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || |: | : :|  ::   |  :  ||  |: :|||:: || ||:| |
Db        4 MLNIAIRAARKAGNVIAKGYERRDDLQTTLKSTNDYVTNIDKASEEAIIEVIRKSYPDHT 63

Qy      136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
            |  ||:|    ::| |  ||:||:||| :|: | | |   ||:   |:  :||:  ||
Db       64 IITEESGALEGKDS-DIQWVIDPLDGTTNFVKGLPHFSVSIAIRVKGRTEVGVVYDPIRN 122

Qy      196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
            | :  | |:   :|   : | :   ||   | |  |           |  : | ::
Db      123 ELFTAVRGEGAKINDMRLRVENKRDLAGTVLTTGFPFKQTRLMPMQFAMMNNLIQ----- 177

Qy      256 CDCYAY----------ALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
             ||  :           :|:| ||   | |:| :|  :   ::  ||| : |: |
Db      178 -DCADFRRMGSAALDLCYVAAGRVDGYFEVGVKAWDIAAGDLIVREAGGLVCDFNG---- 232

Qy      306 WPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                |  || ::|||   |: || |: ::
Db      233 ----GHSYLTSGHLVAAA-PRIVKEILNKIQ 258


RESULT  12
O32889
ID   O32889      PRELIMINARY;      PRT;   255 AA.
AC   O32889;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
     MONOPHOPHATASE (PUTATIVE MONOPHOSPHATASE).
     ML0662 OR MLCB1779.29.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
DR   EMBL; Z98271; CAB11010.1; -.
DR   EMBL; AL583919; CAC30171.1; -.
DR   Leproma; ML0662; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 2.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   255 AA;  27040 MW;  2633DDC72324D545 CRC64;


  Query Match             16.4%;  Score 283;  DB 16;  Length 255;
  Best Local Similarity   31.4%;  Pred. No. 4.6e-15;
  Matches   85;  Conservative   40;  Mismatches  102;  Indels   44;  Gaps    9;

Qy       81 VAQRAADAAGEVLRKYF-RQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
            :|   || |  :    |    : :  | | :||| |||  |  : :|: :  |   | ||
Db        2 LALTLADRADALTSARFGALNLRVDTKPDLTPVTDADRAVEADVRAVLGRERPKDGILGE 61

Qy      140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW- 198
            | |     :     |::|||||||:|: | |:: :|||||  :| | |||:  | |: ||
Db       62 EYGGTITFSGQQ--WIVDPIDGTKNFVRGVPVWASLIALLEDGVPSIGVVSAPALQRRWW 119

Qy      199 --------IGVDG--------KQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAF 242
                    : |||        :   ||   :|  | :  ||  |             | |
Db      120 AARGQGAFVAVDGVPRRLAVSEVADLNSASLSFSSLSGWAQRGL-----------RDRF 167

Qy      243 IRVRNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
            : : : | :|  || |  :| ||| | :|:  |  :  :|  :|  |:  ||| :|   |
Db      168 LELTDAVWRVRAYG-DFLSYCLLAEGAIDVAAEPKVSVWDLAALDIVVREAGGVLTGLDG 226

Qy      302 DKLHWPVTAESRPTSFNVVAAGDARVHKEAL 332
                      |   : ||  : |:|:| |
Db      227 ---------TPGPHGGSAVAT-NGRLHQEVL 247


RESULT  13
O67791
ID   O67791      PRELIMINARY;      PRT;   264 AA.
AC   O67791;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MYO-INOSITOL-1(OR 4)-MONOPHOSPHATASE.
GN   IMP2 OR AQ_1983.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olsen G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
DR   EMBL; AE000766; AAC07753.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Complete proteome.
SQ   SEQUENCE   264 AA;  29336 MW;  10F32D1D3B81B76A CRC64;


  Query Match             16.1%;  Score 279;  DB 16;  Length 264;
  Best Local Similarity   28.5%;  Pred. No. 1e-14;
  Matches   75;  Conservative   54;  Mismatches  106;  Indels   28;  Gaps    7;

Qy       76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
            :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  |||| ||
Db        5 KKYLEVAKIAALAGGQVLKENFGKVKKENIEEKGEKDFVSYVDKTSEERIKEVILKFFPD 64

Qy      134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
            | : ||| |   || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db       65 HEVVGEEMG---AEGSGSEYRWFIDPLDGTKNYINGFPIFAVSVGLVKGEEPIVGAVYLP 121
```

```
Qy    193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRV 245
          : : |  |    :||: | |:    |  | :    |     |        :| |  |
Db    122 YFDKLYWGAKGLGAYVNGKRIKVKDNESLKHAGVVYGFPSRSRRDISIYLNIFKDVFYEV 181

Qy    246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
          :: |  |       ::| |  | ::|  :||:|  : : ::: |||
Db    182 -GSMRRP--GAAAVDLCMVAEGIFDGMMEFEMKPWDITAGLVILKEAGG----------- 227

Qy    306 WPVTAESRPTSFNVVAAGDARVH 328
          |    |   : : ||:  :|
Db    228 -VYTLVGEPFGVSDIIAGNKALH 249


RESULT  14
Q9K4B1
ID   Q9K4B1      PRELIMINARY;      PRT;   266 AA.
AC   Q9K4B1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE MONOPHOSPHATASE.
GN   SC7E4.05C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL359214; CAB94593.1; -.
DR   HSSP; P29218; 1IMF.
DR   InterPro; IPR002106; AA_tRNA_ligase_II.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00629; IMP_1; 1.
SQ   SEQUENCE   266 AA;  29083 MW;  9C2802DB177B243E CRC64;


  Query Match             16.1%;  Score 278.5;  DB 2;  Length 266;
  Best Local Similarity   30.3%;  Pred. No. 1.1e-14;
  Matches   83;  Conservative   51;  Mismatches  111;  Indels   29;  Gaps   10;

Qy     77 RLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAI 136
          ||  |   |||||   : ::    :::  | | :||: ||: ||| :   : :: |  ::
Db      9 RLAHVLADAADAA--TMDRFKALDLKVETKPDMTPVSEADKAAEELIRGHLSRARPRDSV 66

Qy    137 FGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH----NGKPVIGVIDQP 192
           ||| |      :    ||:||||||||::: | |:: |||||:       :|||:|::  |
Db     67 HGEEFG---VAGTGPRRWVIDPIDGTKNYVRGVPVWATLIALMEAKEGGYQPVVGLVSAP 123

Qy    193 ILRERWIGVD------GKQTTLNGQEISVRSCNLLAQA-YLYTTSPHLFEADAEDAFIRV 245
          |  ||  |:      |: | : :  : |   : |: | : |::     |    | |: :
Db    124 ALGRRWWAVEDHGAFTGRSLT-SAHRLHVSQVSTLSDASFAYSSLSGWEEQGRLDGFLDL 182

Qy    246 RNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKL 304
```

```
          :| :   || | : | ::| | ||:  |  |  :|  :   ::  |||: |   |
Db    183 TREVWRTRAYG-DFWPYMMVAEGSVDLCAEPELSLWDMAANAIIVTEAGGTFTGLDG--- 238

Qy    305 HWPVTAESRPTSFNVVAAGDARVHKEALDALRWR 338
                  | | |  || : |:| | |  |  |
Db    239 ------RPGPHSGN-AAASNGRLHDELLGYLNQR 265


RESULT  15
P95189
ID   P95189      PRELIMINARY;      PRT;   260 AA.
AC   P95189;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 27.7 KDA PROTEIN.
GN   RV3137 OR MTCY03A2.21C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; Z83867; CAB06277.1; -.
DR   TubercuList; Rv3137; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 2.
DR   PROSITE; PS00629; IMP_1; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   260 AA;  27993 MW;  CBCD2C2AA3E01518 CRC64;


  Query Match             15.8%;  Score 273.5;  DB 16;  Length 260;
  Best Local Similarity   32.5%;  Pred. No. 2.8e-14;
  Matches   88;  Conservative   42;  Mismatches  104;  Indels   37;  Gaps   11;

Qy     81 VAQRAADAAGEVLR-KYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
          :|   || | |: | ::    : |  | | :||| |||  |  :   : :  |   : ||
Db      8 LALALADRADELTRVRFGALDLRIDTKPDLTPVTDADRAVESDVRQTLGRDRPGDGVLGE 67

Qy    140 ENGWRCAENSADFV---WVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
          | |      |  |    |::|||||||:|: | |:: :|||||  :| | :||:  | |:
Db     68 EFG-----GSTTFTGRQWIVDPIDGTKNFVRGVPVWASLIALLEDGVPSVGVVSAPALQR 122

Qy    197 RWIGVDGKQ--TTLNG---QEISVRSCNLLAQAYLYTTS------PHLFEADAEDAFIRV 245
          ||    |:    :::|     :|| |   |  | | |  :|      | |      : || :
Db    123 RWWAARGRGAFASVDGARPHRLSVSSVAELHSASLSFSSLSGWARPGL-----RERFIGL 177

Qy    246 RNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKL 304
           : | :|  || |  :| |:| | |||  |  :  :|  :|  |:  ||| :|   |
Db    178 TDTVWRVRAYG-DFLSYCLVAEGAVDIAAEPQVSVWDLAALDIVVREAGGRLTSLDG--- 233

Qy    305 HWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
                : |   : ||  :  :| | |  |
Db    234 ------VAGPHGGSAVAT-NGLLHDEVLTRL 257


Search completed: May 30, 2002, 17:16:34
Job time: 210 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 30, 2002, 17:04:58 ; Search time 32.63 Seconds
(without alignments)
1150.565 Million cell updates/sec

Title: US-09-686-522A-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1730	100.0	338	21	AAY44230	Corn extragenic su
2	1294.5	74.8	324	21	AAY44233	Partial wheat extr
3	1067	61.7	295	21	AAY44232	Soybean extragenic
4	310.5	17.9	260	22	AAG90637	C glutamicum prote
5	310.5	17.9	260	22	AAB79381	Corynebacterium gl
6	310.5	17.9	260	22	AAB79382	Corynebacterium gl
7	286	16.5	265	19	AAW42392	Aquifex pyrophilus
8	279	16.1	264	19	AAW42388	Aquifex VF-5 alkal
9	267.5	15.5	271	21	AAG21330	Arabidopsis thalia
10	260	15.0	361	21	AAG11457	Arabidopsis thalia
11	260	15.0	371	21	AAG11456	Arabidopsis thalia
12	260	15.0	400	21	AAG11455	Arabidopsis thalia
13	255.5	14.8	267	20	AAW97883	Maize myo-inositol
14	253.5	14.7	371	21	AAG39265	Arabidopsis thalia
15	253.5	14.7	400	21	AAG39264	Arabidopsis thalia
16	251.5	14.5	361	21	AAG39266	Arabidopsis thalia
17	244.5	14.1	266	22	AAU67151	Propionibacterium
18	237.5	13.7	267	21	AAY44227	Wheat myo-inositol
19	236.5	13.7	233	21	AAB41258	Human ORFX ORF1022
20	224	12.9	339	20	AAW86325	Kidney injury asso
21	221	12.8	168	21	AAY44225	Partial soybean my
22	212	12.3	249	21	AAY44228	Partial barley ext
23	212	12.3	284	22	ABB66131	Drosophila melanog
24	210.5	12.2	273	22	AAG90638	C glutamicum prote
25	210.5	12.2	288	22	ABB66138	Drosophila melanog
26	208.5	12.1	291	22	AAB79387	Corynebacterium gl
27	208.5	12.1	291	22	AAB79388	Corynebacterium gl
28	207	12.0	254	22	AAB96106	Putative inositol
29	207	12.0	397	21	AAG36298	Arabidopsis thalia
30	204	11.8	264	22	AAU55613	Propionibacterium
31	201	11.6	278	22	ABB71757	Drosophila melanog
32	200	11.6	284	22	ABB66140	Drosophila melanog
33	199.5	11.5	270	15	AAR45333	Rat IMP. Rattus r
34	196.5	11.4	288	19	AAW60676	IMP-18p myo-inosit
35	195.5	11.3	253	22	AAG82667	S. epidermidis ope
36	194.5	11.2	596	22	ABB71755	Drosophila melanog
37	193.5	11.2	277	15	AAR45332	Human brain IMP.
38	193.5	11.2	299	22	ABG16258	Novel human diagno
39	188.5	10.9	325	20	AAY35574	Chlamydia pneumoni
40	187.5	10.8	206	22	AAU67657	Propionibacterium
41	185	10.7	406	21	AAG49087	Arabidopsis thalia
42	185	10.7	407	21	AAG49086	Arabidopsis thalia
43	185	10.7	416	21	AAG49085	Arabidopsis thalia
44	181	10.5	406	21	AAG20079	Arabidopsis thalia
45	181	10.5	407	21	AAG20078	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY44230
ID AAY44230 standard; Protein; 338 AA.
XX
AC AAY44230;
XX
DT 21-FEB-2000 (first entry)
XX
DE Corn extragenic suppressor protein.
XX
KW Clone cdt2c.pk003.b20; corn extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW animal feed.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT Misc-difference 11
FT /note= "Encoded by GCC"
FT Misc-difference 94
FT /note= "Encoded by ANG"
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;

XX
DR WPI; 2000-072181/06.
DR N-PSDB; AAZ29165.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 6; Fig 2; 60pp; English.
XX
CC The present sequence is corn extragenic suppressor protein encoded by the
CC cDNA insert in clone cdt2c.pk003.b20. This belongs to the inositol
CC monophosphate family of proteins. The cDNA clone was isolated from corn
CC developing tassle. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising the corresponding cDNA. This is done to increase
CC levels of free phosphate and decrease phytic acid levels in developing
CC seeds or in the construction of transgenic plants with altered levels of
CC the enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. This is
CC commercially used for animal feed.
XX
SQ Sequence 338 AA;

Query Match 100.0%; Score 1730; DB 21; Length 338;
Best Local Similarity 100.0%; Pred. No. 1e-169;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MLSSSSSTHSDTSPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mlssssthsdtspfpglasanpnprsrllrlraaspvssavlsasgrqpmstvrasfaa 60

Qy      61 GAAGRRAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAE 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gaagrraaavgelaterlvevaqraadaagevlrkyfrqrveiidkedhspvtiadreae 120

Qy     121 EAMVSVILKSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 eamvsvilksfpthaifgeengwrcaensadfvwvldpidgtksfitgkplfgtliallh 180

Qy     181 NGKPVIGVIDQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAED 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ngkpvigvidqpilrerwigvdgkqttlngqeisvrscnllaqaylyttsphlfeadaed 240

Qy     241 AFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 afirvrnkvkvplygcdcyayallasgfvdivvesglkpydflslvpviegaggsitdwr 300

       301 GDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALRWR 338
           ||||||||||||||||||||||||||||||||||||
Db     301 gdklhwpvtaesrptsfnvvaagdarvhkealdalrwr 338
```

RESULT 2
AAY44233
ID AAY44233 standard; Protein; 324 AA.
XX
AC AAY44233;
XX
DT 21-FEB-2000 (first entry)
XX
DE Partial wheat extragenic suppressor protein.
XX
KW Clone wlm0.pk0010.f6; wheat extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW animal feed.
XX
OS Triticum aestivum.
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR N-PSDB; AAZ29168.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
PS Claim 6; Fig 2; 60pp; English.
XX
CC The present sequence is a portion of wheat extragenic suppressor protein
CC encoded by the cDNA insert in clone wlm0.pk0010.f6. This belongs to the
CC inositol monophosphate family of proteins. The cDNA clone was isolated
CC from wheat seedlings. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising the corresponding cDNA. This is done to increase
CC levels of free phosphate and decrease phytic acid levels in developing
CC seeds or in the construction of transgenic plants with altered levels of
CC the enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. This is
CC commercially used for animal feed.
XX
SQ Sequence 324 AA;

Query Match 74.8%; Score 1294.5; DB 21; Length 324;
Best Local Similarity 80.6%; Pred. No. 8.1e-125;
Matches 254; Conservative 24; Mismatches 28; Indels 9; Gaps 3;

```
Qy      32 LRAASPVSSAVLSASGR------QPMSTVRASFAAGAAGRRAAAVGE--LATERLVEVAQ 83
           ||: | :::   ||:||      : | :||||  : | |   || |:  :  |||| |||
Db      11 lrspsllatfsssaagracgiagrwmgsvras-pseaggwavaaagkegvdmerlvavaq 69

Qy      84 RAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGEENGW 143
           |||||||||||||| |||||||||||||||||||| ||||||||||:|||||||
Db      70 saadaagevlrkyfrqrfeiidkedhspvtiadreaeeamtsvilksfpthavfgeengw 129

Qy     144 RCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDG 203
           |||| |||:|||||||||||||||||||||||||||||:|:||||||||||:||||
Db     130 rcaeksadyvwvldpidgtksfitgkplfgtliallhngkpvmgiidqpilrerwvgvdg 189

Qy     204 KQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYGCDCYAYAL 263
           |:|||||||||| ||:| |||||||||||| ||||||||||:|||||||||||||||||
Db     190 kkttlngqeisvrpcnvleqaylyttsphlfegdaedafirvrdkvkvplygcdcyayal 249

Qy     264 LASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPTSFNVVAAG 323
           |||||||:|||||||||||||||||||||||||| |:||||||::|||||||||||||
Db     250 lasgfvdlvvesglkpydflslvpviegaggsitdwegnklhwpvssesrptsfnvvaag 309

Qy     324 DARVHKEALDALRWR 338
           |: || :|| |||||
Db     310 dshvhgqalaalrwr 324
```

RESULT 3
AAY44232
ID AAY44232 standard; Protein; 295 AA.
XX
AC AAY44232;
XX
DT 21-FEB-2000 (first entry)
XX
DE Soybean extragenic suppressor protein.
XX
KW Clone s12.pk122.p24; clone src3c.pk013.g15; clone sfl1n.pk003.g19;
KW soybean extragenic suppressor protein; phytic acid biosynthetic enzyme;

KW transformation; chimeric gene; antibody; free phosphate;
KW phytic acid level; transgenic plant; enzyme detection; animal feed.
XX
OS Glycine max.
XX
PN WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF 22-APR-1999; 99WO-US08791.
XX
PR 24-APR-1998; 98US-0082960.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
DR WPI; 2000-072181/06.
DR N-PSDB; AAZ29167.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the enzyme in transgenic plants -
XX
PS Claim 6; Fig 2; 60pp; English.
XX
CC The present sequence is a soybean extragenic suppressor protein encoded
CC by a contig assembled from portions of the cDNA inserts in clones
CC s12.pk122.p24, src3c.pk013.g15, and sfl1n.pk003.g19. This belongs to the
CC inositol monophosphate family of proteins. The level of expression
CC of this phytic acid biosynthetic enzyme can be altered by transforming a
CC host cell with the chimeric gene comprising the corresponding cDNA. This
CC is done to increase levels of free phosphate and decrease phytic acid
CC levels in developing seeds or in the construction of transgenic plants
CC with altered levels of the enzyme. The phytic acid biosynthetic enzymes
CC are used to raise antibodies, for enzyme detection in cells and cell
CC extracts. This is commercially used for animal feed.
XX
SQ Sequence 295 AA;

Query Match 61.7%; Score 1067; DB 21; Length 295;
Best Local Similarity 63.0%; Pred. No. 2.1e-101;
Matches 201; Conservative 42; Mismatches 44; Indels 32; Gaps 2;

```
Qy      20 SANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGELATERLV 79
           | :| | :   |||| :| |:                              |   |
Db       9 shspipnt-tfrlramaphst------------------------------plelnrfa 36

Qy      80 EVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           ||  : |||||||:|||||:  ::| | | ||||||| :|||||||:|| :|||:|||:||
Db      37 evgnkvadaagevirkyfrknfdvihkhdlspvtiadqsaeeamvsiildnfpshaiyge 96

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWI 199
           |||||| | :||:|||||||||||||||||:||||:||| || |::|:||||:||||||
Db      97 engwrceeknadyvwvldpidgtksfitgkpvfgtlvallqngtpilgiidqpvlrerwi 156

Qy     200 GVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYGCDCY 259
           |: ||:|:||||||| |:|  |:|||||||||||  |||:||||||:|||  |||||||
Db     157 giagkrtslngqeistrtcadlsqaylyttsphlfngdaeeafirvrskvkfqlygcdcy 216

Qy     260 AYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPTSFNV 319
           |||||:|||||:||||||||||||:|:|||||||| ||||:|||| |  :  |  |||||
Db     217 ayallssgfvdlvvesglkpydflalipviegaggvitdwkgdklfweasplsiatsfnv 276

Qy     320 VAAGDARVHKEALDALRWR 338
           ||||| ::|::|||:|:|:
Db     277 vaagdkqihqqaldslqwk 295
```

RESULT 4
AAG90637
ID AAG90637 standard; Protein; 260 AA.
XX
AC AAG90637;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 4391.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH65856.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 4391; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 260 AA;

Query Match 17.9%; Score 310.5; DB 22; Length 260;
Best Local Similarity 33.3%; Pred. No. 1.8e-23;
Matches 91; Conservative 42; Mismatches 101; Indels 39; Gaps 11;

```
Qy      81 VAQRAADAAGEV-LRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: |  : | ::    :|:  | | :||: ||   |||:   |  : |  :| ||
Db      10 lalelaeladsitldrfeasdlevsskpdmtpvsdadlateealrekiatarpadsilge 69

Qy     140 ENGWRCAENSADF---VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
           | |        :|    |::||||||||::: | |:: |||||| ||||| |||  | |
Db      70 efg-----gdvefsgrqwiidpidgtknyvrgvpvwatlialldngkpvagvisapalar 124

Qy     197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTTSPHLFEADAEDAFIRVRNKV- 249
           ||   :|     | ||   :::||   : |  | | :::     | |  | |: : :
Db     125 rwwasegagawrtfngssprklsvsqvskldaslsfsslsgwaerdlrdqfvsltdttw 184

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG-DKLHWPV 308
           ::  || | ::| |:| | |||  |  :  :|   |  ::  |||  |   | |  |
Db     185 rlrgyg-dffsyclvaegavdiaaepevslwdlaplsilvteaggkftslagvdgph--- 240
```

```
Qy     309 TAESRPTSFNVVAAGDA-----RVHKEALDALR 336
                     |||      :| | || |:
Db     241 ------------ggdavatngilhdetldrlk 260


RESULT   5
AAB79381
ID   AAB79381 standard; Protein; 260 AA.
XX
AC   AAB79381;
XX
DT   30-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum SMP protein sequence SEQ ID NO:278.
XX
KW   Corynebacterium glutamicum; carbon metabolism and energy production;
KW   SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW   fine chemical production; organic acid; proteinogenic amino acid;
KW   nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW   nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
     carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
     diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100844-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00943.
XX
PR   25-JUN-1999;   99US-0141031.
PR   08-JUL-1999;   99DE-1031412.
PR   08-JUL-1999;   99DE-1031413.
PR   08-JUL-1999;   99DE-1031419.
PR   08-JUL-1999;   99DE-1031420.
PR   08-JUL-1999;   99DE-1031424.
PR   08-JUL-1999;   99DE-1031428.
PR   08-JUL-1999;   99DE-1031431.
PR   08-JUL-1999;   99DE-1031433.
PR   08-JUL-1999;   99DE-1031434.
PR   08-JUL-1999;   99DE-1031510.
PR   08-JUL-1999;   99DE-1031562.
PR   08-JUL-1999;   99DE-1031634.
PR   09-JUL-1999;   99DE-1032180.
PR   09-JUL-1999;   99DE-1032227.
PR   09-JUL-1999;   99DE-1032230.
PR   09-JUL-1999;   99US-0143208.
     14-JUL-1999;   99DE-1032924.
     14-JUL-1999;   99DE-1032973.
     14-JUL-1999;   99DE-1033005.
PR   27-AUG-1999;   99DE-1040765.
PR   31-AUG-1999;   99US-0151572.
PR   03-SEP-1999;   99DE-1042076.
PR   03-SEP-1999;   99DE-1042079.
PR   03-SEP-1999;   99DE-1042086.
PR   03-SEP-1999;   99DE-1042087.
PR   03-SEP-1999;   99DE-1042088.
PR   03-SEP-1999;   99DE-1042095.
PR   03-SEP-1999;   99DE-1042123.
PR   03-SEP-1999;   99DE-1042125.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-061975/07.
DR   N-PSDB; AAF71498.
XX
PT   New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT   metabolism and oxidative phosphorylation protein for production or
PT   modulation of production of fine chemicals e.g. amino acids,
PT   carbohydrates or enzymes -
XX
PS   Claim 20; Page 537-538; 1246pp; English.
XX
CC   AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC   metabolism and oxidative phosphorylation (SMP) proteins given in
CC   AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC   energy production. The C. glutamicum SMP gene can be used in vectors
CC   (II) for expression in host cells and production or modulation of
CC   production of fine chemicals, such as, an organic acid, a proteinogenic
CC   or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC   a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC   acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC   cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC   (III) encoded by them are used for diagnosing the presence or activity of
CC   Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC   containing them are used to map genomes of organisms related to
CC   C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC   in evolutionary studies, in determining SMP protein regions required
CC   for function, in modulating SMP protein activity, in modulating the
CC   metabolism of sugars, and in modulating high-energy molecule production
CC   in a cell (i.e. ATP, NADPH).
XX
SQ   Sequence   260 AA;


  Query Match             17.9%;  Score 310.5;  DB 22;  Length 260;
  Best Local Similarity   33.3%;  Pred. No. 1.8e-23;
  Matches   91;  Conservative   42;  Mismatches  101;  Indels   39;  Gaps   11;

Qy      81 VAQRAADAAGEV-LRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: |  : | ::    :|:  | | :||: ||   |||:   |  : |  :| ||
Db      10 lalelaeladsitldrfeasdlevsskpdmtpvsdadlateealrekiatarpadsilge 69

Qy     140 ENGWRCAENSADF---VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
           | |        :|    |::||||||||::: | |:: |||||| ||||| |||  | |
Db      70 efg-----gdvefsgrqwiidpidgtknyvrgvpvwatlialldngkpvagvisapalar 124

Qy     197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTTSPHLFEADAEDAFIRVRNKV- 249
           ||   :|     | ||   :::||   : |  | | :::     | |  | |: : :
Db     125 rwwasegagawrtfngssprklsvsqvskldda slsfsslsgwaerdlrdqfvsltdttw 184

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG-DKLHWPV 308
           ::  || | ::| |:| | ||| |  :  :|   |  ::  |||  |   | |  |
Db     185 rlrgyg-dffsyclvaegavdiaaepevslwdlaplsilvteaggkftslagvdgph--- 240

Qy     309 TAESRPTSFNVVAAGDA--   RVHKEALDALR 336
                     |||      :| | || |:
Db     241 ------------ggdavatngilhdetldrlk 260


RESULT   6
AAB79382
ID   AAB79382 standard; Protein; 260 AA.
XX
AC   AAB79382;
XX
DT   30-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum SMP protein sequence SEQ ID NO:280.
XX
KW   Corynebacterium glutamicum; carbon metabolism and energy production;
KW   SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW   fine chemical production; organic acid; proteinogenic amino acid;
KW   nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW   nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW   carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW   diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100844-A2.
XX
PD   04-JAN-2001.
```

XX
PF 23-JUN-2000; 2000WO-IB00943.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031412.
PR 08-JUL-1999; 99DE-1031413.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031431.
PR 08-JUL-1999; 99DE-1031433.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031562.
PR 08-JUL-1999; 99DE-1031634.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-061975/07.
DR N-PSDB; AAF71499.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 20; Page 540; 1246pp; English.
XX
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 260 AA;

Query Match 17.9%; Score 310.5; DB 22; Length 260;
Best Local Similarity 33.3%; Pred. No. 1.8e-23;
Matches 91; Conservative 42; Mismatches 101; Indels 39; Gaps 11;

```
Qy      81 VAQRAADAAGEV-LRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   |: |  : | ::    :|:  | | :||: ||   |||:   |  : |  :| ||
Db      10 lalelaeladsitldrfeasdlevsskpdmtpvsdadlateealrekiatarpadsilge 69

Qy     140 ENGWRCAENSADF---VWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
           | |        :|    |::|||||||::: | |:: |||||| ||||| |||  | |
Db      70 efg-----gdvefsgrqwiidpidgtknyvrgvpvwatlialldngkpvagvisapalar 124

Qy     197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTTSPHLFEADAEDAFIRVRNKV- 249
           ||   :|     | ||   :::||   : |  | | :::     | |  | |: : :
Db     125 rwwasegagawrtfngssprklsvsqvskldd aslsfsslsgwaerdlrdqfvsltdttw 184

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG-DKLHWPV 308
           ::  || | ::| |:| | |||  |  :  :|   |  ::  |||  |   | |  |
Db     185 rlrgyg-dffsyclvaegavdiaaepevslwdlaplsilvteaggkftslagvdgph--- 240

Qy     309 TAESRPTSFNVVAAGDA-----RVHKEALDALR 336
                       |||      :| | || |:
Db     241 ------------ggdavatngilhdetldrlk 260
```

RESULT 7
AAW42392
ID AAW42392 standard; Protein; 265 AA.
XX
AC AAW42392;
XX
DT 22-JUN-1998 (first entry)
XX
DE Aquifex pyrophilus phosphatase (28ph1).
XX
KW Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW food; detergent; baking.
XX
OS Aquifex pyrophilus.
XX
PN WO9748416-A1.
XX
PD 24-DEC-1997.
XX
PF 19-JUN-1997; 97WO-US10784.
XX
PR 19-JUN-1996; 96US-0033752.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Bylina E, Lee E, Mathur EJ;
XX
DR WPI; 1998-062851/06.
DR N-PSDB; AAV03317.
XX
PT Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT detergent, and baking industries
XX
PS Claim 11; Page 95; 128pp; English.
XX
CC This protein comprises a thermostable phosphatase, designated
CC 28ph1, of Aquifex pyrophilus. The invention relates to claimed
CC polynucleotides (see AAV03301-20) coding for claimed thermostable
CC phosphatases (see AAW42380-95). Vector and host cells are used to
CC produce the enzymes, which can be used in a claimed method to
CC hydrolyse phosphate bonds. They can also be used in enzyme
CC labelling processes, in certain recombinant DNA techniques, in
CC ELISA immunoassays, in enzyme linked gene probes, in research
CC applications for removing 5' phosphates in polynucleotides prior to
CC end labelling, and in the pharmaceutical, food, detergent, and
CC baking industries.
XX
SQ Sequence 265 AA;

Query Match 16.5%; Score 286; DB 19; Length 265;

```
  Best Local Similarity   29.8%;  Pred. No. 6.4e-21;
  Matches   78;  Conservative   53;  Mismatches  105;  Indels   26;  Gaps    8;

Qy      76 ERLVEVAQRAADAAGEVLRKYF-RQRVE-IIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           |:|:|||: || | |:||:: | : ::| | :| :   |:  |: :|| :  :||| ||
Db       5 ekllevakmaalaggqvlkenfgkiklenieekgekdfvsyvdktseerikelilkffpd 64

Qy     134 HAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPI 193
           | : ||| |    |  : : | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 hevvgeergkegke--spykwfidpldgtknyikgfpifavsvglvkenepivgavylpy 122

Qy     194 LRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRVR 246
             :   |:    ||: |||:    |  | :    |      |          :: |  |
Db     123 fdtlywaskgrgaykngerisvkergelkhaavvygfpsrsrrdislylnvfkevfyev- 181

Qy     247 NKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHW 306
             |: |  |       :|| |  | ::|  :||:|  : : ::: |||  |  :||
Db     182 gsvrrp--gaaavdicmlaegifdgmmefemkpwditaglvilkeaggfyt-lkgd---- 234

Qy     307 PVTAESRPTSFNVVAAGDARVH 328
               |  : : ||:  :|
       235 -------pfgisdiiagnrmlh 249


RESULT   8
AAW42388
ID   AAW42388 standard; Protein; 264 AA.
XX
AC   AAW42388;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Aquifex VF-5 alkaline phosphatase (34A1A).
XX
KW   Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW   food; detergent; baking.
XX
OS   Aquifex sp. strain VF-5.
XX
PN   WO9748416-A1.
XX
PD   24-DEC-1997.
XX
PF   19-JUN-1997;   97WO-US10784.
XX
PR   19-JUN-1996;   96US-0033752.
XX
PA   (RECO-) RECOMBINANT BIOCATALYSIS INC.

     Bylina E,  Lee E,  Mathur EJ;
XX
DR   WPI; 1998-062851/06.
DR   N-PSDB; AAV03309.
XX
PT   Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT   detergent, and baking industries
XX
PS   Claim 10; Fig 9; 128pp; English.
XX
CC   This protein comprises a mature thermostable alkaline phosphatase
CC   (34A1A) of Aquifex VF5, a marine bacterium that grows optimally
CC   at 85-90 degC and pH 6.8.  Its amino acid sequence, deduced from
CC   an isolated polynucleotide (see AAV03309), shows 57% identity to
CC   Escherichia coli suppressor protein suhB.  The invention relates to
CC   claimed polynucleotides (see AAV03301-09 and AAV03310-20) coding for
CC   claimed thermostable phosphatases (see AAW42380-95).  Vector and host
CC   cells are used to produce the enzymes, which can be used in a
CC   claimed method to hydrolyse phosphate bonds.  They can also be used
CC   in enzyme labelling processes, in recombinant DNA techniques, in
CC   ELISA, in enzyme linked gene probes, in research applications for
CC   removing 5' phosphates in polynucleotides prior to end labelling,
CC   and in the pharmaceutical, food, detergent, and baking industries.
XX

SQ   Sequence   264 AA;


  Query Match             16.1%;  Score 279;  DB 19;  Length 264;
  Best Local Similarity   28.5%;  Pred. No. 3.3e-20;
  Matches   75;  Conservative   54;  Mismatches  106;  Indels   28;  Gaps    7;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  ||||  ||
Db       5 kkylevakiaalaggqvlkenfgkvkkenieekgekdfvsyvdktseerikevilkffpd 64

Qy     134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           | : ||| |    || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 hevvgeemg---aegsgseyrwfidpldgtknyingfpifavsvglvkgeepivgavylp 121

Qy     193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRV 245
             : : |  |     :|| | |:    |  | :    |     |          :| |  |
Db     122 yfdklywgakglgayvngkrikvkdneslkhagvvygfpsrsrrdisiylnifkdvfyev 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
             :: |  |       ::| |  | ::|  :||:|  : : ::: |||
Db     182 -gsmrrp--gaaavdlcmvaegifdgmmefemkpwditaglvilkeagg---------- 227

Qy     306 WPVTAESRPTSFNVVAAGDARVH 328
             |    |   : : ||:  :|
Db     228 -vytlvgepfgvsdiiagnkalh 249


RESULT   9
AAG21330
ID   AAG21330 standard; Protein; 271 AA.
XX
AC   AAG21330;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 23850.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
```

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

```
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             15.5%;  Score 267.5;  DB 21;  Length 271;
  Best Local Similarity   30.1%;  Pred. No. 5.4e-19;
  Matches   81;  Conservative   50;  Mismatches  121;  Indels   17;  Gaps    7;

Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :  |  ||  ||:::|| | :   :  |     ||  |:  || : : : : || |
Db       8 dqflaaaidaakkagqiirkgfyetkhvehksqvdlvtetdkgceelvfnhlkqlfpnhk 67

       136 IFGEEN--GWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPI 193
             |||    :   | : :  |::||:||| :|: | |     | |     ||:||:  ||
Db      68 figeettaafgvteltdeptwivdpldgttnfvhgfpfvcvsigltigkvpvvgvvynpi 127

Qy     194 LRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT-TSPHLFEADAEDAFIRVRN---KV 249
           : | : || ||   |||: | | : : |  | | |       :|  :|   |: :   ||
Db     128 meelftgvqgkgaflngkrikvsaqselltallvteagtkrdkatlddttnrinslltkv 187

Qy     250 K-VPLYGCDCYAYALLASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWRGDKLHWP 307
           : : : |        :| | |||  | |   |:|  : : ::: ||| | |  |  |
Db     188 rslrmsgscaldlcgvacgrvdifyelgfggpwdiaaqivivkeagglifdpsgkdl--d 245

Qy     308 VTAESRPTSFNVVAAGDARVHKEALDALR 336
           :|::        :|| :| : :   :|||
Db     246 itsqr-------iaasnaslkelfaealr 267


RESULT  10
AAG11457
ID   AAG11457 standard; Protein; 361 AA.
XX
AC   AAG11457;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 10179.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
     hybridisation assay; genetic mapping; gene expression control; promoter;
     termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
```

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 260; DB 21; Length 361;
Best Local Similarity 26.9%; Pred. No. 4.9e-18;
Matches 92; Conservative 54; Mismatches 152; Indels 44; Gaps 10;

```
Qy      13 SPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGE 72
           :|| |  : |   | |  |: : |  |:  |    :  :| |          |  |   |
Db      17 npfsg-rtvnrtfryrctrilsnsfksttrlqt--kavlsevsdqtryprig--akttgt 71

Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           ::   |:|| : ||    ||: :   :   |  |     ||  |: :| |:: |: |:|
Db      72 ispahllevvelaaktgaevvmeavnkprnitykglsdlvtdtdkaseaailevvknfs 131

Qy     133 THAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKP----VIGV 188
           | | ||| |    ::|:|::| :||:||| :|  | | |   : :|: | |    |:
Db     132 dhlilgeegg-iigdsssdylwcidpldgttnfahgypsfavsvgvlyrgnpaaasvvef 190

Qy     189 IDQPILR--ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT----------TSPHLFE 235
           :  |: |    :    |     |||:| |   : : :| | |          |:  ||:
Db     191 vggpmcrntrtfsatagggalcngqkihvsktdaverallitgfgyehddawstnmelfk 250

Qy     236 ADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGS 295
               |  |   |:    |        :| |  :   |  |||:|  : | ::| |||:
Db     251 -----eftdvsrgvr--rlgaaavdmchvalgiaesyweyrlkpwdmaagvliveeagga 303

Qy     296 ITDWRGDK---------  --LHWPVTAESRPTSFNVVAAG 323
           :|   | |            ||  :     | : |: : |
Db     304 vtrmdggkfsvidrsvlvsngvlhpkllerlapatenlkskg 345
```

RESULT 11
AAG11456
ID AAG11456 standard; Protein; 371 AA.
XX
AC AAG11456;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10178.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 260; DB 21; Length 371;
Best Local Similarity 26.9%; Pred. No. 5.1e-18;
Matches 92; Conservative 54; Mismatches 152; Indels 44; Gaps 10;

```
Qy      13 SPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGE 72
           :|| |  : |   | |  |: : |  |:  |    :  :| |         |  |   |
Db      27 npfsg-rtvnrtfryrctrilsnsfksttrlqt--kavlsevsdqtryprig--akttgt 81

Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           ::   |:|| : ||    ||: :   :  |  |     ||  |: :| |:: |: |:|
Db      82 ispahllevvelaaktgaevvmeavnkprnitykglsdlvtdtdkaseaailevvkknfs 141

Qy     133 THAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKP----VIGV 188
            | | ||| |    ::|:|::| :||:||| :|  | | |   : :|: | |    |:
Db     142 dhlilgeegg-iigdsssdylwcidpldgttnfahgypsfavsvgvlyrgnpaaasvvef 200

Qy     189 IDQPILR--ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT----------TSPHLFE 235
           :  |: |    :    |     |||:| |   : : :| | |          |:  ||:
Db     201 vggpmcrntrtfsatagggalcngqkihvsktdaverallitgfgyehddawstnmelfk 260

Qy     236 ADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGS 295
                |  |   |:   |        :| |  :   |  |||:|  : | ::| |||:
Db     261 -----eftdvsrgvr--rlgaaavdmchvalgiaesyweyrlkpwdmaagvliveeagga 313

Qy     296 ITDWRGD----------LHWPVTAESRPTSFVVAAG 325
           :|   | |          ||  :    | : |: : |
Db     314 vtrmdggksvfdrsvlvsngvlhpkllerіapatenlkskg 355
```

RESULT 12
AAG11455
ID AAG11455 standard; Protein; 400 AA.
XX
AC AAG11455;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10177.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 260; DB 21; Length 400;
Best Local Similarity 26.9%; Pred. No. 5.7e-18;
Matches 92; Conservative 54; Mismatches 152; Indels 44; Gaps 10;

```
Qy      13 SPFPGLASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAAVGE 72
           :|| |  : |   | |  |: : |  |:  |    :  :| |        |  |   |
Db      56 npfsg-rtvnrtfryrctrilsnsfksttrlqt--kavlsevsdqtryprig--akttgt 110

Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           ::   |:|| : ||    ||: :   :   |  |     ||  |: :| |:: |: |:|
Db     111 ispahllevvelaaktgaevvmeavnkprnitykglsdlvtdtdkaseaailevvkknfs 170

Qy     133 THAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKP----VIGV 188
            | | ||| |    ::|:|::| :||:||| :|  | | |   : :|: | |    |:
Db     171 dhlilgeegg-iigdsssdylwcidpldgttnfahgypsfavsvgvlyrgnpaaasvvef 229

Qy     189 IDQPILR--ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT----------TSPHLFE 235
           :  |: |    :    |     |||:| |   : : :| | |          |:  ||:
Db     230 vggpmcrntrtfsatagggalcngqkihvsktdaverallitgfgyehddawstnmelfk 289

Qy     236 ADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGS 295
               |  |   |:    |        :| |  :   |  |||:|  : | ::| |||:
Db     290 -----eftdvsrgvr--rlgaaavdmchvalgiaesyweyrlkpwdmaagvliveeagga 342

Qy     296 ITDWRGDK-------------LHWPVTAESRPTSFNVVAAG 323
           :|   | |             ||  :    | : |: : |
```

```
Db    343 vtrmdggkfsvfdrsvlvsngvlhpkllleriapatenlkskg 384


RESULT  13
AAW97883
ID   AAW97883 standard; Protein; 267 AA.
XX
AC   AAW97883;
XX
DT   07-JUN-1999  (first entry)
XX
DE   Maize myo-inositol monophosphatase-3.
XX
KW   Myo-inositol monophosphatase-3; maize; corn; phytate;
KW   phytic acid; transgenic plant; animal nutrition; feedstuff; food.
XX
OS   Zea mays.
XX
PN   WO9905298-A1.
XX
PD   04-FEB-1999.
XX
PF   17-JUL-1998;   98WO-US14657.
XX
PR   18-MAY-1998;   98US-0085852.
PR   22-JUL-1997;   97US-0053371.
PR   28-JUL-1997;   97US-0053944.
PR   08-AUG-1997;   97US-0055526.
PR   11-AUG-1997;   97US-0055446.
XX
PA   (PION-) PIONEER HI-BRED INT INC.
XX
PI   Beach LR,  Bowen BA,  Martino-Catt SJ,  Wang H,  Wang X;
XX
DR   WPI; 1999-142948/12.
DR   N-PSDB; AAX24412.
XX
PT   New polynucleotides controlling phytate metabolism in plants -
PT   useful for improving the nutritional content of plants, by enhancing
PT   levels of non-phytate phosphorus, and reducing phytate levels
XX
PS   Claim 1a; Page 81-82; 86pp; English.
XX
CC   This is the amino acid sequence of maize myo-inositol
CC   monophosphatase-3, an enzyme involved in the metabolism of
CC   phytate.  cDNA (see AAX24410) encoding the enzyme was isolated from
CC   a maize immature ear cDNA library.  Polynucleotides (see AAX24400,
CC   AAX24403, AAX24407 and AAX24410-12) encoding maize
CC   phosphatidylinositol-3-kinase (see AAW97880), myo-inositol
CC   1,3,4-triphosphate 5/6-kinase (see AAW97881), myo-inositol
CC   1-phosphate synthase and myo-inositol monophosphatase-3, all
CC   enzymes involved in phytate metabolism, are claimed.  The
CC   invention relates to the use of such genes to reduce the
CC   levels of phytate, and/or increase the levels of non-phytate
CC   phosphorus, in plants used for food or feed.  The genes are
CC   especially used to improve the nutritional content of plants such
CC   as corn and soybean.  Transgenic plants, and seed produced by them,
CC   are claimed.
XX
SQ   Sequence   267 AA;


  Query Match             14.8%;  Score 255.5;  DB 20;  Length 267;
  Best Local Similarity   30.3%;  Pred. No. 9e-18;
  Matches   73;  Conservative   41;  Mismatches  112;  Indels   15;  Gaps    5;

Qy     73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
          :: |: : ||  || :|||::|| | |   :  |     ||  |:  |: : : : | ||
Db      1 mseeqflavaveaaksageiirkgfyqtknvqhkgqvdlvtetdkacedlifnhlrkhfp 60

Qy    133 THAIFGEENG---WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           |   |||       |: : |  |::||:||| :|: | |     : |     | :||:
Db     61 dhkfigeeesaalgatadltddptwivdpldgttnfvhgfpfvcvsvqltigkiptvgvv 120

Qy    190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
            ||: | :  | ||   |||  |   | : | :| | | :    :    |      |::
Db    121 fnpimnelftavrgkgaflngspikassqdelvkallvteagtnrdkttvd---dttnri 177

Qy    250 KVPLYGCD----CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
            ||       | : ||    :| | :|:  | |   |:|  :   ::: ||| : |
Db    178 nrllykirsirmcgslalnmcgvacgrldlcyeigfggpwdvaagavilqeagglvfdps 237

Qy    301 G 301
           |
Db    238 g 238


RESULT  14
AAG39265
ID   AAG39265 standard; Protein; 371 AA.
XX
AC   AAG39265;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 48558.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-01  825.
PR   05-MAR-1999;   99US-0[illegible] 180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
```

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157965.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159647.
PR 14-OCT-1999; 99US-0159648.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.7%; Score 253.5; DB 21; Length 371;
Best Local Similarity 25.8%; Pred. No. 2.4e-17;

```
  Matches   93;  Conservative   55;  Mismatches  165;  Indels   47;  Gaps   10;

Qy       1 MLSSSSSTHSDTSPFPGLASANP---NPRSRLLRLRAASPVSSAVLSASGRQP---MSTV 54
           : | : |      |   |   ||      :|  | |    :|::  | :  |    :| |
Db       6 ifsgnmslrishlprsslplqnpisgrtvnrtfryrctrilsnsfksttrlqtkavlsev 65

Qy      55 RASFAAGAAGRRAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTI 114
                 |  |   | ::   |:|| : ||    ||: :   :   |  |     ||
Db      66 sdqtryprig--akttgtispahllevvelaaktgaevvmnavnkprnitykglsdlvtd 123

Qy     115 ADREAEEAMVSVILKSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGT 174
            |: :| |:: |: |:|  | | ||| |    ::|:|::| :||:||| :|  | | |
Db     124 tdkaseaailevvkknfsdhlilgeegg-iigdsssdylwcidpldgttnfahgypsfav 182

Qy     175 LIALLHNGKP----VIGVIDQPIL--RERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT 228
            : :|: | |    |:  :  |:      :    |     |||:| |   : : :| | |
Db     183 svgvlyrgnpaaasvvefvggpmcwntrtfsatagggalcngqkihvsktdaverallit 242

Qy     229 ----------TSPHLFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGL 277
                    |:  ||:      |  |   |:    |        :| |  :   |  |
.o     243 gfgyehddawstnmelfk-----eftdvsrgvr--rlgaaavdmchvalgiaesyweyrl 295

Qy     278 KPYDFLSLVPVIEGAGGSITDWRGDK-------------LHWPVTAESRPTSFNVVAAG 323
           ||:|  : | ::| |||::|   | |              ||  :     | : |: : |
Db     296 kpwdmaagvliveeaggavtrmdggkfsvfdrsvlvsngvlhpklleriapatenlkskg 355


RESULT  15
AAG39264
ID   AAG39264 standard; Protein; 400 AA.
XX
AC   AAG39264;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 48557.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
```

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
29-SEP-1999; 99US-0156596.
04-OCT-1999; 99US-0157117.
R 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.7%; Score 253.5; DB 21; Length 400;
Best Local Similarity 25.8%; Pred. No. 2.7e-17;
Matches 93; Conservative 55; Mismatches 165; Indels 47; Gaps 10;

```
Qy       1 MLSSSSSTHSDTSPFPGLASANP---NPRSRLLRLRAASPVSSAVLSASGRQP---MSTV 54
           : | : |       |   |   ||       :|  | |    :|::  | :  |    :| |
Db      35 ifsgnmslrishlprsslplqnpisgrtvnrtfryrctrilsnsfksttrlqtkavlsev 94

Qy      55 RASFAAGAAGRRAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTI 114
                 |  |   | ::   |:|| : ||    ||: :   :   |  |     ||
Db      95 sdqtryprig--akttgtispahllevvelaaktgaevvmeavnkprnitykglsdlvtd 152

Qy     115 ADREAEEAMVSVILKSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGT 174
             |: :| |:: |: |:|  | | ||| |    ::|:|::| :||:||| :|  | | |
Db     153 tdkaseaailevvkknfsdhlilgeegg-iigdsssdylwcidpldgttnfahgypsfav 211

Qy     175 LIALLHNGKP----VIGVIDQPIL--RERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT 228
             : :|: | |    |:  :  |:      :    |     |||:| |   : : :| | |
Db     212 svgvlyrgnpaaasvvefvggpmcwntrtfsatagggalcngqkihvsktdaverallit 271

Qy     229 ----------TSPHLFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGL 277
                     |:  ||:      |  |   |:    |        :| |  :   |  |
Db     272 gfgyehddawstnmelfk-----eftdvsrgvr--rlgaaavdmchvalgiaesyweyrl 324

Qy     278 KPYDFLSLVPVIEGAGGSITDWRGDK-------------LHWPVTAESRPTSFNVVAAG 323
           ||:|  : | ::| |||::|   | |              ||  :     | : |: : |
Db     325 kpwdmaagvliveeaggavtrmdggkfsvfdrsvlvsngvlhpkllerіapatenlkskg 384
```

Search completed: May 30, 2002, 17:15:12
Job time: 614 sec

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 17, 2002, 16:19:08 ; Search time 45 Seconds
(without alignments)
722.076 Million cell updates/sec

Title: US-09-686-522C-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1020	59.0	286	2	T08560	hypothetical prote
2	525	30.3	262	2	AG3034	inositol monophosp
3	525	30.3	297	2	E98251	similar to mlcb177
4	483	27.9	308	2	F95293	probable inositol
5	477	27.6	278	2	C87528	inositol monophosp
6	463.5	26.8	272	2	AF3515	extragenic suppres
7	396.5	22.9	267	2	S74899	extragenic suppres
8	293	16.9	267	2	AC0353	probable inositol
9	292	16.9	271	2	C83169	extragenic suppres
10	283	16.4	255	2	T45317	monophophatase [im
11	279	16.1	264	1	C70470	myo-inositol-1(or
12	273.5	15.8	260	2	B70646	probable monophosp
13	272	15.7	267	1	F64103	suppressor protein
14	265.5	15.3	268	2	T07795	inositol-1(or 4)-m
15	258	14.9	267	1	D65030	inositol-1(or 4)-m
16	258	14.9	267	2	G91053	extragenic suppres
17	257	14.9	267	2	AI0824	extragenic suppres
18	254.5	14.7	270	2	AF2170	inositol monophoph
19	248	14.3	288	2	D82285	inositol monophosp
20	245	14.2	266	2	AG3580	inositol-1(or 4)-m
21	244.5	14.1	269	2	AG3145	inositol monophosp
22	244.5	14.1	276	2	D98142	spcA protein (AF17
23	244	14.1	263	2	AH3292	inositol-1(or 4)-m
24	240	13.9	266	2	H98266	extragenic suppres
25	240	13.9	266	2	AH3017	extragenic suppres
26	239.5	13.8	266	2	E84292	extragenic suppres
27	238	13.8	265	2	T07800	inositol-1(or 4)-m
28	236.5	13.7	268	2	F97443	inositol monophosp
29	236.5	13.7	268	2	AH2661	inositol monophosp
30	235	13.6	270	2	T12205	inositol-1(or 4)-m
31	234	13.5	275	2	E82554	extragenic supress
32	231.5	13.4	353	2	A86438	hypothetical prote
33	230	13.3	287	1	S76736	hypothetical prote
34	229.5	13.3	256	2	T50642	myo-inositol-1(or
35	229.5	13.3	275	2	AE3393	inositol monophosp
36	226.5	13.1	232	2	E72255	hypothetical prote
37	221.5	12.8	265	1	E69864	myo-inositol-1(or
38	221	12.8	266	2	D95885	probable inositol
39	221	12.8	272	2	T03436	accG protein - Agr
40	221	12.8	272	2	AG3245	arabinose phosphat
41	221	12.8	280	2	AB2875	exopolysaccharide
42	220	12.7	262	2	C97651	pssB protein [impo
43	220	12.7	294	2	G95861	probable inositol
44	218	12.6	257	2	AB1208	extragenic suppres
45	218	12.6	304	2	G97583	inositol monophosp

ALIGNMENTS

RESULT 1
T08560
hypothetical protein T22F8.20 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000
C;Accession: T08560
R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16442
A;Accession: T08560
A;Molecule type: DNA
A;Residues: 1-286 <BEV>
A;Cross-references: EMBL:AL050351; GSPDB:GN00062; ATSP:T22F8.20
A;Experimental source: cultivar Columbia; BAC clone T22F8
C;Genetics:
A;Gene: ATSP:T22F8.20
A;Map position: 4
A;Introns: 58/1; 84/2; 116/1; 132/3; 174/2; 196/3; 226/3; 264/1
C;Superfamily: suppressor protein suhB

Query Match 59.0%; Score 1020; DB 2; Length 286;
Best Local Similarity 69.5%; Pred. No. 9.3e-73;
Matches 182; Conservative 45; Mismatches 35; Indels 0; Gaps 0;

```
Qy      76 ERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :|    |    |||:|||:|||||:: :|:||:| |||||||: |||||||:| :: |:||
Db      24 DRFAAVGNALADASGEVIRKYFRKKFDIVDKDDMSPVTIADQMAEEAMVSIIFQNLPSHA 83

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |:||| |||| | |||:||||||||||||||||||:||||||||: |||::|:||||||:
Db      84 IYGEEKGWRCKEESADYVWVLDPIDGTKSFITGKPVFGTLIALLYKGKPILGLIDQPILK 143

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYG 255
           ||||||::|::| |||::|| ||| |:||||||||||||  :|| |: |||:||||||||
Db     144 ERWIGMNGRRTKLNGEDISTRSCPKLSQAYLYTTSPHLFSEEAEKAYSRVRDKVKVPLYG 203

Qy     256 CDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRPT 315
           |||||||||||||||:|:|||||||||:|||||||||||:|||| | :  |  :: :  |
Db     204 CDCYAYALLASGFVDLVIESGLKPYDFLALVPVIEGAGGTITDWTGKRFLWEASSSAVAT 263

Qy     316 SFNVVAAGDARVHKEALDALRW 337
           |||||||||: :|::||::| |
Db     264 SFNVVAAGDSDIHQQALESLEW 285
```

RESULT 2
AG3034
inositol monophosphatase family protein [imported] - Agrobacterium tumefaciens (stra
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AG3034
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wo

```
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AG3034
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44693.1; PID:g17742322; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3885
A;Map position: linear chromosome
C;Superfamily: suppressor protein suhB

  Query Match             30.3%;  Score 525;  DB 2;  Length 262;
  Best Local Similarity   42.7%;  Pred. No. 7.2e-34;
  Matches  112;  Conservative   51;  Mismatches   73;  Indels   26;  Gaps    7;

Qy      84 RAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
           : ||||       ||  : :|:|:|  : |||  |: || |: ::| : || | | |||:
Db      14 KLADAASAETLPRFRTGIAVINKQDGGYDPVTEGDQAAETAIRALIEERFPQHGILGEEH 73

Qy     142 GWRCAEN---SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
           |     |     | :||:||||||::||:| |::||||    :|:  :|::|||  :||:
Db      74 G-----NVGLDRDHIWVIDPIDGTRAFISGVPVWGTLIGFQSSGRATMGIMDQPFTKERY 128

Qy     199 IGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
           |||     |    ::|  | |  |: | |:||:||:| |: :  : :|:::|::  |
Db     129 F-ADGKAAWYFGPDGEKKIRTRDCASLSDAVLFTTTPHIFTAEEKPLYEKVQDQVRLFRY 187

Qy     255 GCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRP 314
           | ||||| |||:| ||:|:|||||||  :|:|||| ||'::| | |           ||
Db     188 GVDCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAG TMTTWDG----------GRP 237

Qy     315 -TSFNVVAAGDARVHKEALDAL 335
            ::|||   ||:|||  |
Db     238 ENGGRILAAGSKAVHEEALAIL 259


RESULT 3
E98251
similar to mlcb1779.29 (probable monophosphatase gene) in mycobacterium leprae [importe
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 15-Feb-2002
C;Accession: E98251
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: E98251
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-297 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89535.1; PID:g15159414; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1926
A;Map position: linear chromosome
C;Superfamily: suppressor protein suhB

  Query Match             30.3%;  Score 525;  DB 2;  Length 297;
  Best Local Similarity   42.7%;  Pred. No. 8.4e-34;
  Matches  112;  Conservative   51;  Mismatches   73;  Indels   26;  Gaps    7;

Qy      84 RAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
           : ||||       ||  : :|:|:|  : |||  |: || |: ::| : || | | |||:
Db      49 KLADAASAETLPRFRTGIAVINKQDGGYDPVTEGDQAAETAIRALIEERFPQHGILGEEH 108

Qy     142 GWRCAEN---SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW 198
           |     |     | :||:||||||::||:| |::||||    :|:  :|::|||  :||:
Db     109 G-----NVGLDRDHIWVIDPIDGTRAFISGVPVWGTLIGFQSSGRATMGIMDQPFTKERY 163

Qy     199 IGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254
           |||     |    ::|  | |  |: | |:||:||:| |: :  : :|:::|::  |
Db     164 F-ADGKAAWYFGPDGEKKIRTRDCASLSDAVLFTTTPHIFTAEEKPLYEKVQDQVRLFRY 222

Qy     255 GCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTAESRP 314
           | ||||| |||:| ||:|:|||||||  :|:|||| ||'::| | |           ||
Db     223 GVDCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAGGTMTTWDG----------GRP 272

Qy     315 -TSFNVVAAGDARVHKEALDAL 335
            ::|||   ||:|||  |
Db     273 ENGGRILAAGSKAVHEEALAIL 294


RESULT 4
F95293
probable inositol monophosphatase [imported] - Sinorhizobium meliloti (strain 1021)
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: F95293
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.;
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh,
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium mel
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: F95293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-308 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK64912.1; PID:g14523333; GSPDB:GN00165
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hub
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R
 L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lela
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: SMa0483
A;Genome: plasmid

  Query Match             27.9%;  Score 483;  DB 2;  Length 308;
  Best Local Similarity   39.6%;  Pred. No. 1.8e-30;
  Matches  107;  Conservative   38;  Mismatches  103;  Indels   22;  Gaps    4;

Qy      70 VGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILK 129
           :|| |:       |   || | : :       | | | | | ||||  ||  |: :   |
Db      56 LGEFAS-----FAHDIADLARQTISSAAGVRREPIAKSDASPVTETDRAVEKCLRERIAD 110

Qy     130 SFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           || | : ||| |     |  :||||:|||||||:|: | |::||||:|   | |::|:|
Db     111 HFPDHGVLGEEFGAEGLGN--EFVWVIDPIDGTKAFVAGLPVYGTLISLTRGGTPILGLI 168

Qy     190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
           | |:  :||:|| |: ||||   |   |   || |::   :|  |    :     :|
Db     169 DNPMTGDRWLGVSGQPTTLNNVPIRTASTTALATAFIANGNPDAFSPADKSRVESLRRIT 228

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
           :  :|| | ||  :| | ||| :: || |||: :|||| |||| ||||:|
Db     229 RWCVYGGSCIAYGRVADGSVDISIDGGLDPYDYCALVPVITGAGGCITDWQG-------- 280

Qy     310 AESRPTSFN----VVAAGDARVHKEALDAL 335
           || : |     ||     :|:  |: |
Db     281 ---RPLTLNSGGLCVATATDLLHRHVLEIL 307


RESULT 5
C87528
```

```
inositol monophosphatase family protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 01-Mar-2002
C;Accession: C87528
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87528
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-278 <STO>
A;Cross-references: GB:AE005673; NID:g13423761; PIDN:AAK24223.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2252
C;Superfamily: suppressor protein suhB

  Query Match             27.6%;  Score 477;  DB 2;  Length 278;
  Best Local Similarity   40.8%;  Pred. No. 4.6e-30;
  Matches  116;  Conservative   48;  Mismatches   88;  Indels   32;  Gaps    9;

Qy      73 LATERLVEVAQRAAD---AAGEVLRKYFRQRVEIIDK--------EDHS---PVTIADRE 118
           |: :||  :     |   |: :|:   ||    : ||        : |:   ||| |||
Db       3 LSADRLAALDAFILDLNRASADVILPLFRADHGLEDKGAGKNLPRDTHAAFDPVTEADRG 62

Qy     119 AEEAMVSVILKSFPTHAIFGEENGWRCAEN--SADFVWVLDPIDGTKSFITGKPLFGTLI 176
           || |: ::| : || | : ||| |    |:  |:|||||||||||::|| | ||: |||
Db      63 AEAAIRALIAQRFPDHGVIGEEYG----EDRPDAEFVWVLDPIDGTRAFIA PLWTTLI 118

Qy     177 ALLHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQE---ISVRSCNLLAQAYLYTTSPH- 232
           | | |:||:| | || : | :||  |    ::| |   | || |  :  | : || |
Db     119 GLRHEGRPVLGSIGQPYVNEIFIGHAGGARLVSGGEARPIRVRECANINDAVIATTDPDA 178

Qy     233 LFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGA 292
           |:     |:::||   |:   ||| ||||::| | :|:|:|:|||| :|  : :|:||||
Db     179 CFDGAERGAWLQVRAAAKLARLGCDAYAYAMVAMGKMDMV EAGLKSWDILAAIPLIEGA 238

Qy     293 GGSITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           || :|:|||:    ||     |    :| :|| |   ||| :||
Db     239 GGMVTNWRGE----PV----GPNGGQMVISGDRRPLDEALVSLR 274


RESULT 6
AF3515
extragenic suppressor protein suhb [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 ;Accession: AF3515
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AF3515
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-272 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL53289.1; PID:g17984172; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEII0048
A;Map position: II
C;Superfamily: suppressor protein suhB

  Query Match             26.8%;  Score 463.5;  DB 2;  Length 272;
  Best Local Similarity   41.3%;  Pred. No. 5.2e-29;
  Matches  116;  Conservative   47;  Mismatches   93;  Indels   25;  Gaps    9;

Qy      66 RAAAVGELATERLVEVA---QRAADAAGEVLRKYFRQRVEIIDKED--HSPVTIADREAE 120
           ||   |   |  |:: |   : || || : | : |||  |: :|      ||| ||| ||
Db       4 RAPCFGNRRTGLLIDKAFFSEVAAAAAAQTLPR-FRQLTEVDNKYSVGFDPVTEADRAAE 62

Qy     121 EAMVSVILKSFPTHAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALL 179
           |: :|| ::|| | | ||| |   |||:    ||::||:|||::||:| |::|||: |
Db      63 RAIRAVIGRTFPYHGILGEEYG---AENTDRSHVWIIDPVDGTRAFISGLPVWGTLVGLT 119

Qy     180 HNGKPVIGVIDQPILRERWIGVDGKQTTLN-----GQEISVRSCNLLAQAYLYTTSPHLF 234
           :|    |:: ||   | :   ||    |       : : ||   :|  | |:||:| ||
Db     120 VDGDARAGMMSQPFTGELFYS-DGDGAYLQRGDGAPRRLCVRKNAVLEDATLFTTTPALF 178

Qy     235 EADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGG 294
           : |   || |: : |::  || ||||:|:|| ||||||||:||: ||  :|:|:|| |||
Db     179 KGDDRKAFDRLESAVRLSRYGVDCYAFAMLAGGFVDIVVEAGLQTYDIAALIPIIEQAGG 238

Qy     295 SITDWRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           :|   |         ||      ::|||    :|: ||| |
Db     239 VVTRRDGG------PAEQ---GGDIVAAATPALHQAALDLL 270


RESULT 7
S74899
extragenic suppressor - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein sll1959
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S74899
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Y
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechoc
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74899
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <KAN>
A;Cross-references: EMBL:D90909; GB:AB001339; NID:g1652844; PIDN:BAA17860.1; PID:gl
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: suppressor protein suhB

  Query Match             22.9%;  Score 396.5;  DB 2;  Length 267;
  Best Local Similarity   36.9%;  Pred. No. 9.4e-24;
  Matches  101;  Conservative   49;  Mismatches   99;  Indels   25;  Gaps   10;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR----QRVEIIDKEDHSPVTIADREAEEAMVSVILKSF 131
           |: : :||: |  :||:| :|||     |    ||:   : || || |||:||| :|    |
Db       6 EQRLFIAQQLAAVSGEILIQYFRRSHLQGGTKIDQVS-AIVTQADEEAEQAMVDLIQAQF 64

Qy     132 PTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLH-NGKPVIGVID 190
           |    :  || |    |  |  : |||||||||| ||: | |:| ||| |:  : :||:|:
Db      65 PQDGVIREE-GKNIAGKSG-YTWVLDPIDGTSSFVRGLPIFATLIGLVDADMRPVLGIAH 122

Qy     191 QPILRERWIGVDGKQTTLNG----QEISVRSCNLLAQAYLYTTSPHLFEADA-EDAFIRV 245
           |||  :|| || |:|: :||             || | | : :|:| :|     :     :
Db     123 QPISGDRWQGVQGEQSNVNGIPLVNPYKASEINLTA-ACIVSTTPLMFTTPVQQQKMADI 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVD---IVVESGLKPYDFLSLVPVIEGAGGSITDWRGD 302
           : :   :| ||: |   |||:     ::||: |  ||| :|:|:: ||    |||:|
Db     182 YRQCQRTAFGGDCFNYLSAASGWTAMPLVIVEADLNFYDFCALIPILTGANYCFTDWQGK 241

Qy     303 KLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
           :|         | |  |||: : ::| | |  |:
Db     242 EL--------TPESTEVVASPNPKLHSEILAFLQ 267


RESULT 8
AC0353
probable inositol monophosphatase family protein YPO2899 [imported] - Yersinia pest
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AC0353
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,
```

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0353
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC92150.1; PID:g15980865; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2899
C;Superfamily: suppressor protein suhB

Query Match 16.9%; Score 293; DB 2; Length 267;
Best Local Similarity 31.6%; Pred. No. 1.3e-15;
Matches 86; Conservative 43; Mismatches 111; Indels 32; Gaps 8;

```
Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| ||| || :: | :    ||   |  :  ||  ||:||  :: || ||:| |
Db       4 MLTIAIRAARKAGNLIAKNYETPDAVEASQKGSNDFVTNVDRDAEHLIIDVIRKSYPKHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |    |:  |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db      64 IISEECGELVGEDD-DVQWVIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCN-------LLAQAYLYTTSPHLFEADAEDAFIRVRNK 248
           | :    |:   |||  :  |  |       :||  : :    |        |:|||  |
Db     123 ELFTASRGQGAQLNGYRL--RGTNAKDLDGTILATGFPFKVKQH------APAYIRVVGK 174

Qy     249 V-----KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           :          |       | :|:| ||   | ||||:||     ::  :|| :|: |
Db     175 LFEQCADFRRTGSAALDLAYVAAGRVDGFFEIGLKPWDFAGGELLVRESGGIVTDFAGGH 234

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
            |:        :| |:| ||: |: |  : |:
Db     235 NHF--------SSGNIV-AGNPRIVKSIVQAM 257
```

RESULT 9
C83169
extragenic suppressor protein SuhB PA3818 [imported] - Pseudomonas aeruginosa (strain PA
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: C83169
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: C83169
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-271 <STO>
A;Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIDN:AAG07205.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3818
C;Superfamily: suppressor protein suhB

Query Match 16.9%; Score 292; DB 2; Length 271;
Best Local Similarity 29.7%; Pred. No. 1.6e-15;
Matches 82; Conservative 54; Mismatches 106; Indels 34; Gaps 8;

```
Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEII---DKEDHSPVTIADREAEEAMVSVILKSFP 132
           :: :| ||| :|||::   ||  :|:::|   :|:    ||  || ||: :|: : |::|
Db       4 MLNIALRAARSAGELI---FRSIERLDVISVNEKDAKDYVTEVDRAAEQTIVAALRKAYP 60

Qy     133 THAIFGEENGW-RCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
           |||| ||| |:   :   ||::||:||:||| :|| | | |   ||  : |:    |:
Db      61 THAIMGEEGGFIEGSGEGADYLWVIDPLDGTTNFIHGVPHFAVSIACKYKGRLEHAVVLD 120

Qy     192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHL-FEADAEDAFIR-----V 245
```

```
           |: :| :    |:   |||: : |     |  | | |  |    : |  | ::      |
Db     121 PVRQEEFTASRGRGAALNGRRLRVSGRKSLEGALLGTGFPFRDNQIDNLDNYLNMFRSLV 180

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG---- 301
               :   |       | :|:|  |   | ||  :|  :   ::: ||| ::|: |
Db     181 GQTAGIRRAGAASLDLAYVAAGRYDAFWEFGLSEWDMAAGALLVQEAGGLVSDFTGSHEF 240

Qy     302 -DKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
            :| |              : ||: :  |  |  ::
Db     241 LEKGH-------------IVAGNTKCFKALLTTIQ 262
```

RESULT 10
T45317
monophophatase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 18-Feb-2000
C;Accession: T45317
R;Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z22864
A;Accession: T45317
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-255 <PAR>
A;Cross-references: EMBL:Z98271; PIDN:CAB11010.1
A;Experimental source: cosmid B1779
C;Genetics:
A;Note: MLCB1779.29
C;Superfamily: suppressor protein suhB

Query Match 16.4%; Score 283; DB 2; Length 255;
Best Local Similarity 31.4%; Pred. No. 7.5e-15;
Matches 85; Conservative 40; Mismatches 102; Indels 44; Gaps 9;

```
Qy      81 VAQRAADAAGEVLRKYF-RQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   || |  :    |    : :  | | :||| |||  |  : :|: :  |   | ||
Db       2 LALTLADRADALTSARFGALNLRVDTKPDLTPVTDADRAVEADVRAVLGRERPKDGILGE 61

Qy     140 ENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERW- 198
           | |     :     |::|||||||:|: | |:: :|||| :| | |||:  | |: ||
Db      62 EYGGTITFSGQQ--WIVDPIDGTKNFVRGVPVWASLIALLEDGVPSIGVVSAPALQRRWW 119

Qy     199 --------IGVDG--------KQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAF 242
                 : |||         :   ||   :|  | :  ||  |             | |
Db     120 AARGQGAFVAVDGVPRRLAVSEVADLNSASLSFSSLSGWAQRGL------------RDRF 167

Qy     243 IRVRNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
           : : : | :|  || |  :| ||| | :|:  |  :  :|  :|  |:  ||| :|   |
Db     168 LELTDAVWRVKAYG-DFLSYCLLAEGAIDVAAEPKVSVWDLAALDIVVREAGGVLTGLDG 226

Qy     302 DKLHWPVTAESRPTSFNVVAAGDARVHKEAL 332
                   |   : ||  : |:|:| |
Db     227 ---------TPGPHGGSAVAT-NGRLHQEVL 247
```

RESULT 11
C70470
myo-inositol-1(or 4)-monophosphatase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: C70470
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: C70470
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-264 <AQF>
A;Cross-references: GB:AE000766; GB:AE000657; NID:g2984216; PIDN:AAC07753.1; PID:g29
A;Experimental source: strain VF5

```
C;Genetics:
A;Gene: imp2
C;Superfamily: suppressor protein suhB

  Query Match           16.1%;  Score 279;  DB 1;  Length 264;
  Best Local Similarity  28.5%;  Pred. No. 1.6e-14;
  Matches   75;  Conservative   54;  Mismatches  106;  Indels   28;  Gaps    7;

Qy      76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  ||||  ||
Db       5 KKYLEVAKIAALAGGQVLKENFGKVKKENIEEKGEKDFVSYVDKTSEERIKEVILKFFPD 64

Qy     134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           | : ||| |   || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 HEVVGEEMG---AEGSGSEYRWFIDPLDGTKNYINGFPIFAVSVGLVKGEEPIVGAVYLP 121

Qy     193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRV 245
              : |  |    :||: | |:    |  | :    |        :| |  |
Db     122 YFDKLYWGAKGLGAYVNGKRIKVKDNESLKHAGVVYGFPSRSRRDISIYLNIFKDVFYEV 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
           :: |  |       ::| |  | ::|  :||:|  : : ::: |||
Db     182 -GSMRRP--GAAAVDLCMVAEGIFDGMMEFEMKPWDITALVILKEAGG---------- 227

Qy     306 WPVTAESRPTSFNVVAAGDARVH 328
           |    |   : : ||:  :|
Db     228 -VYTLVGEPFGVSDIIAGNKALH 249


RESULT 12
B70646
probable monophosphatase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: B70646
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70646
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-260 <COL>
A;Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CAB06277.1; PID:g1781230
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3137
C;Superfamily: suppressor protein suhB

  Query Match           15.8%;  Score 273.5;  DB 2;  Length 260;
  Best Local Similarity  32.5%;  Pred. No. 4.3e-14;
  Matches   88;  Conservative   42;  Mismatches  104;  Indels   37;  Gaps   11;

Qy      81 VAQRAADAAGEVLR-KYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGE 139
           :|   || | |: | ::    : |  | | :||| |||  |  :   : :  |   : ||
Db       8 LALALADRADELTRVRFGALDLRIDTKPDLTPVTDADRAVESDVRQTLGRDRPGDGVLGE 67

Qy     140 ENGWRCAENSADFV---WVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
           | |      |  |    |::|||||||:|: | |:: :|||||  :| | :||:  | |:
Db      68 EFG-----GSTTFTGRQWIVDPIDGTKNFVRGVPVWASLIALLEDGVPSVGVVSAPALQR 122

Qy     197 RWIGVDGKQ--TTLNG---QEISVRSCNLLAQAYLYTTS------PHLFEADAEDAFIRV 245
           ||    |:    :::|     :|| |   |  | | :|      | |      : || :
Db     123 RWWAARGRGAFASVDGARPHRLSVSSVAELHSASLSFSSLSGWARPGL-----RERFIGL 177

Qy     246 RNKV-KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKL 304
            : | :|  || |  :| |:| | |||  |  :  :|  :|  |:  ||| :|   |
Db     178 TDTVWRVRAYG-DFLSYCLVAEGAVDIAAEPQVSVWDLAALDIVVREAGGRLTSLDG--- 233

Qy     305 HWPVTAESRPTSFNVVAAGDARVHKEALDAL 335


                  : |   : ||  :  :| | |  |
Db     234 ------VAGPHGGSAVAT-NGLLHDEVLTRL 257


RESULT 13
F64103
suppressor protein suhB - Haemophilus influenzae (strain Rd KW20)
C;Species: Haemophilus influenzae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C;Accession: F64103
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavag
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidma
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vent
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: F64103
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-267 <TIGR>
A;Cross-references: GB:U32775; GB:L42023; NID:g1573951; PIDN:AAC22595.1; PID:g15739
C;Genetics:
A;Gene: suhB
C;Superfamily: suppressor protein suhB

  Query Match           15.7%;  Score 272;  DB 1;  Length 267;
  Best Local Similarity  30.8%;  Pred. No. 5.8e-14;
  Matches   84;  Conservative   44;  Mismatches  113;  Indels   32;  Gaps    8;

Qy      78 LVEVAQRAADAAGEVLRKYFRQR--VEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || |: | : :|  :|   |  :  ||  |: :|  :: || ||:| |
Db       4 MLNIAIRAARKAGNVIAKNYERRDAIESTQKGINDYVTNVDKASEAEIIEVIRKSYPDHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILR 195
           |  || |        :|  |::||:|||::|:|| | |   ||:    :  :||:  ||
Db      64 IITEETG-AIEGKDSDVQWIIDPLDGTRNFMTGLPHFSVSIAVRVKNRTEVGVVYDPIRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPH----------LFEADAEDAF-I 243
           | :  | |:   ||    · | |   :  : | |  |           :  |  |||
Db     123 ELFTAVRGEGAKLNEVRI VDSKREIQGSILATGFPFKQPKLMPAQFAMMNALIEDAADF 182

Qy     244 RVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303
           |       : |    ||    :||  :|   | ||| :|  :   ::  ||| : |:
Db     183 RRTGSAALDL----CY----VASNRIDGYFEMGLKAWDCAAGDLIVREAGGLVCDFD--- 231

Qy     304 LHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
                   |   | |::||  :|| || |: :|
Db     232 -----AGNSYLRSGNIIAA PSRVIKEMLNKIR 258


RESULT 14
T07795
inositol-1(or 4)-monophosphatase (EC 3.1.3.25) isoform 3 - tomato
N;Alternate names: myo-inositol-1-phosphatase
C;Species: Lycopersicon esculentum (tomato)
C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 03-Jun-2002
C;Accession: T07795
R;Gillaspy, G.E.; Keddie, J.S.; Oda, K.; Gruissem, W.
Plant Cell 7, 2175-2185, 1995
A;Title: Plant inositol monophosphatase is a lithium-sensitive enzyme encoded by a m
A;Reference number: Z16138; MUID:96351935; PMID:8718627
A;Accession: T07795
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-268 <GIL>
A;Cross-references: EMBL:U39059; NID:g1098970; PIDN:AAB19031.1; PID:g1098971
A;Experimental source: strain VFNT Cherry
C;Genetics:
A;Gene: IMP3
C;Function:
A;Description: removes phosphate from myo-inositol-1-phosphate and myo- inositol 4-p
A;Pathway: inositol phosphate metabolism
```

```
C;Superfamily: suppressor protein suhB
C;Keywords: phosphoric monoester hydrolase

  Query Match             15.3%;  Score 265.5;  DB 2;  Length 268;
  Best Local Similarity   30.2%;  Pred. No. 1.9e-13;
  Matches   83;  Conservative   48;  Mismatches  117;  Indels   27;  Gaps    8;

Qy      74 ATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           : |: ::|| || |||::|: | :   :  |     || |: |: : : : : ||:
Db       6 SVEQFLDVAVEAAKKAGEIIREGFYKTKHVEHKGMVDLVTETDKACEDFIFNHLKQRFPS 65

Qy     134 HAIFGEENGWRCA--ENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQ 191
           |   |||    |   | : :  |::||:|||:|: | |     | |    || :||:
Db      66 HKFIGEETTAACGNFELTDEPTWIVDPLDGTTNFVHGFPFVCVSIGLTIEKKPTVGVVYN 125

Qy     192 PILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYT---TSPHLFEADA-----EDAFI 243
           ||: | : |:|||   |||: | | | : | :| | |   |:      ||
Db     126 PIIDELFTGIDGKGAFLNGKPIKVSSQSELVKALLATEAGTNRDKLVVDATTGRINSLLF 185

(    ,  244 RVRNKVKVPLYGCDCYAYAL--LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
           :|||:    |  |   |  |  :| | :|:  |     |:|      ::: ||| : |
Db     186 KVRS-----LRMCGSCALNLCGVACGRLDLFYELEFGGPWDVAGGAVIVKEAGGFVFDPS 240

Qy     301 GDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
           | :  : :||         |||  :| :    : ||
Db     241 GSE--FDLTARR-------VAATNAHLKDAFIKAL 266


RESULT 15
D65030
inositol-1(or 4)-monophosphatase (EC 3.1.3.25) [validated] - Escherichia coli (strain K-
N;Alternate names: suppressor protein suhB
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
C;Accession: D65030; A35158
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: D65030
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-267 <BLAT>
A;Cross-references: GB:AE000339; GB:U00096; NID:g1788870; PIDN:AAC75586.1; PID:g1788882;
A;Experimental source: strain K-12, substrain MG1655
R;Yano, R.; Nagai, H.; Shiba, K.; Yura, T.
J. Bacteriol. 172, 2124-2130, 1990
A;Title: A mutation that enhances synthesis of sigma(32) and suppresses temperature-sens
A;Reference number: A35158; MUID:90202735; PMID:2136605
A;Accession: A35158
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-140,'L',142-267 <YAN>
A;Cross-references: GB:M34828; NID:g147898; PIDN:AAA67506.1; PID:g147899
C;Genetics:
A;Gene: suhB
C;Complex: monomer [validated, MUID:20213266]
C;Function:
A;Description: EC 3.1.3.25 [validated, MUID:20213266]
A;Note: inositol-1-phosphatase activity is neither sufficient nor required for complemen
C;Superfamily: suppressor protein suhB
C;Keywords: phosphoric monoester hydrolase

  Query Match             14.9%;  Score 258;  DB 1;  Length 267;
  Best Local Similarity   29.6%;  Pred. No. 7.4e-13;
  Matches   79;  Conservative   44;  Mismatches  124;  Indels   20;  Gaps    7;

Qy      78 LVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHA 135
           :: :| |||  || :: | :     ||   |  :  ||  |: ||  ::  | ||:| |
Db       4 MLNIAVRAARKAGNLIAKNYETPDAVEASQKGSNDFVTNVDKAAEAVIIDTIRKSYPQHT 63

Qy     136 IFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGT IALLHNGKPVIGVIDQPILR 195

                 | ||:|       |  ||:||:||| :||   | |   ||:   |:  : |:  |:
Db      64 IITEESG-ELEGTDQDVC..VIDPLDGTTNFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122

Qy     196 ERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEA-DAEDAFIRVRNKV----- 249
           | :    |:  |||  :   :   |    | |  |  |:|      :| :  |:
Db     123 ELFTATRGQGAQLNGYRLRGSTARDLDGTILATGFP--FKAKQYATTYINIVGKLFNECA 180

Qy     250 KVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVT 309
                 |       | :|:| ||   | ||:|:|| :   ::  ||| ::|: |   :: :|
Db     181 DFRRTGSAALDLAYVAAGRVDGFFEIGLRPWDFAAGELLVREAGGIVSDFTGGH-NYMLT 239

Qy     310 AESRPTSFNVVAAGDARVHKEALDALR 336
                   : ||: || |  |  :|
Db     240 GN--------IVAGNPRVVKAMLANMR 258


Search completed: November 1 , 2002, 16:42:29
Job time : 49 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 17, 2002, 16:39:23 ; Search time 20 Seconds
(without alignments)
254.529 Million cell updates/sec

Title: US-09-686-522C-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

otal number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	286	16.5	265	9	US-09-902-525-48	Sequence 48, Appl
2	279	16.1	264	9	US-09-902-525-36	Sequence 36, Appl
3	255.5	14.8	267	10	US-09-921-232-17	Sequence 17, Appl
4	255.5	14.8	267	10	US-09-921-330-17	Sequence 17, Appl
5	255.5	14.8	267	10	US-09-921-329-17	Sequence 17, Appl
6	127	7.3	252	10	US-09-987-446-2	Sequence 2, Appli
7	101	5.8	1616	9	US-09-712-363-262	Sequence 262, App
8	93	5.4	268	10	US-09-216-393-90	Sequence 90, Appl
9	90	5.2	667	10	US-09-896-852-55	Sequence 55, Appl
10	90	5.2	777	10	US-09-925-301-1115	Sequence 1115, Ap
11	88	5.1	3472	9	US-10-027-806-4	Sequence 4, Appli
12	86	5.0	912	10	US-09-817-464-2	Sequence 2, Appli
13	84	4.9	3739	9	US-09-860-846-33	Sequence 33, Appl
14	84	4.9	3739	10	US-09-861-289-33	Sequence 33, Appl
15	84	4.9	11877	9	US-09-860-846-6	Sequence 6, Appli
16	84	4.9	11877	10	US-09-861-289-6	Sequence 6, Appli
17	83.5	4.8	340	10	US-09-159-469-54	Sequence 54, Appl
18	83.5	4.8	340	10	US-09-798-042-54	Sequence 54, Appl
19	83.5	4.8	2383	10	US-09-912-020-302	Sequence 302, App
20	83	4.8	312	10	US-09-815-242-5213	Sequence 5213, Ap
21	83	4.8	499	10	US-09-815-242-11025	Sequence 11025, A
22	82.5	4.8	598	10	US-09-815-242-13230	Sequence 13230, A
23	82.5	4.8	792	10	US-09-815-242-11966	Sequence 11966, A
24	82.5	4.8	943	10	US-09-815-242-11834	Sequence 11834, A
25	82.5	4.8	4613	9	US-09-860-846-31	Sequence 31, Appl
26	82.5	4.8	4613	10	US-09-861-289-31	Sequence 31, Appl
27	82	4.7	799	10	US-09-952-677-6	Sequence 6, Appli
28	82	4.7	953	10	US-09-912-917-2	Sequence 2, Appli
29	81.5	4.7	943	10	US-09-815-242-12027	Sequence 12027, A
30	80.5	4.7	710	10	US-09-801-368-276	Sequence 276, App
31	80	4.6	663	10	US-09-815-242-11869	Sequence 11869, A
32	79.5	4.6	509	9	US-10-121-032-18	Sequence 18, Appl
33	79	4.6	766	10	US-09-925-301-1276	Sequence 1276, Ap
34	79	4.6	820	10	US-09-815-242-10771	Sequence 10771, A
35	78.5	4.5	409	10	US-09-815-242-11270	Sequence 11270, A
36	78.5	4.5	639	10	US-09-854-731-17	Sequence 17, Appl
37	78	4.5	437	10	US-09-815-242-11187	Sequence 11187, A
38	78	4.5	681	10	US-09-815-242-11830	Sequence 11830, A
39	78	4.5	2785	10	US-09-801-574-8	Sequence 8, Appli
40	77	4.5	496	10	US-09-945-301-4	Sequence 4, Appli
41	76.5	4.4	1114	12	US-10-005-983-2	Sequence 2, Appli
42	76	4.4	329	10	US-09-925-300-1406	Sequence 1406, Ap
43	76	4.4	446	12	US-10-081-859-2	Sequence 2, Appli
44	75.5	4.4	280	9	US-09-960-631A-5	Sequence 5, Appli
45	75.5	4.4	294	10	US-09-801-368-202	Sequence 202, App

ALIGNMENTS

```
RESULT 1
US-09-902-525-48
; Sequence 48, Application US/09902525
; Patent No. US20020164751A1
;   GENERAL INFORMATION:
;        APPLICANT: Recombinant Biocatalysis, Inc.
;        TITLE OF INVENTION: THERMOSTABLE PHOSPHATASES
;        NUMBER OF SEQUENCES  54
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fish & Richardson P.C.
;             STREET: 4225 Executive Square, Suite 1400
;             CITY: La Jolla
;             STATE: CA
;             COUNTRY: USA
;             ZIP: 92037
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: Windows 95
;             SOFTWARE: FastSEQ for Windows Version 2.0b
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/902,525
;             FILING DATE: 09-Jul-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/09/202,681A
;             FILING DATE: 23-Dec-1999
;             APPLICATION NUMBER: 08/666,857
;             FILING DATE: 19-JUN-1996
;             APPLICATION NUMBER: 60/033,752
;             FILING DATE: 19-JUN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Haile, Ph.D., Lisa A.
;             REGISTRATION NUMBER: 38,347
;             REFERENCE/DOCKET NUMBER: 09010/015WO1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 619/678-5070
;             TELEFAX: 619/678-5099
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 48:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 265 amino acids
```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-09-902-525-48

Query Match 16.5%; Score 286; DB 9; Length 265;
Best Local Similarity 29.8%; Pred. No. 1.3e-19;
Matches 78; Conservative 53; Mismatches 105; Indels 26; Gaps 8;

```
Qy      76 ERLVEVAQRAADAAGEVLRKYF-RQRVE-IIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           |:|:|||: || | |:||:: | : ::| | :| :   |:  |: :|| :  :||| ||
Db       5 EKLLEVAKMAALAGGQVLKENFGKIKLENIEEKGEKDFVSYVDKTSEERIKELILKFFPD 64

Qy     134 HAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPI 193
           | : ||| |    |  : : | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 HEVVGEERGKEGKE--SPYKWFIDPLDGTKNYIKGFPIFAVSVGLVKENEPIVGAVYLPY 122

v      194 LRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRVR 246
           :    |:    ||: |||:    | | :    |     |         :: |  |
o      123 FDTLYWASKGRGAYKNGERISVKERGELKHAAVVYGFPSRSRRDISLYLNVFKEVFYEV- 181

Qy     247 NKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHW 306
             |: |  |       :|| |  | ::|  :||:|  : : ::: |||  |  :||
Db     182 GSVRRP--GAAAVDICMLAEGIFDGMMEFEMKPWDITAGLVILKEAGGFYT-LKGD---- 234

Qy     307 PVTAESRPTSFNVVAAGDARVH 328
                 |   : : ||:  :|
Db     235 -------PFGISDIIAGNRMLH 249
```

RESULT 2
US-09-902-525-36
; Sequence 36, Application US/09902525
; Patent No. US20020164751A1
; GENERAL INFORMATION:
; APPLICANT: Recombinant Biocatalysis, Inc.
; TITLE OF INVENTION: THERMOSTABLE PHOSPHATASES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/902,525
; FILING DATE: 09-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/202,681A
; FILING DATE: 23-Dec-1999
; APPLICATION NUMBER: 08/666,857
; FILING DATE: 19-JUN-1996
; APPLICATION NUMBER: 60/033,752
; FILING DATE: 19-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/015WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-09-902-525-36

Query Match 16.1%; Score 279; DB 9; Length 264;
Best Local Similarity 28.5%; Pred. No. 6e-19;
Matches 75; Conservative 54; Mismatches 106; Indels 28; Gaps 7;

```
Qy      76 ERLVEVAQRAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPT 133
           :: :|||: || | |:||:: |   ::  | :| :   |:  |: :|| :  ||||  ||
Db       5 KKYLEVAKIAALAGGQVLKENFGKVKKENIEEKGEKDFVSYVDKTSEERIKEVILKFFPD 64

Qy     134 HAIFGEENGWRCAENS-ADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQP 192
           | : ||| |   || | ::: | :||:||||::| | |:|   : |:   :|::| :  |
Db      65 HEVVGEEMG---AEGSGSEYRWFIDPLDGTKNYINGFPIFAVSVGLVKGEEPIVGAVYLP 121

Qy     193 ILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADA-------EDAFIRV 245
           : : |  |    :||: | |:    |  | :    |     |         :| |  |
Db     122 YFDKLYWGAKGLGAYVNGKRIKVKDNESLKHAGVVYGFPSRSRRDISIYLNIFKDVFYEV 181

Qy     246 RNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
           :: |  |       ::| |  | ::|  :||:|  : : ::: |||
Db     182 -GSMRRP--GAAAVDLCMVAEGIFDGMMEFEMKPWDITAGLVILKEAGG---------- 227

Qy     306 WPVTAESRPTSFNVVAAGDARVH 328
           |    |   : : ||:  :|
Db     228 -VYTLVGEPFGVSDIIAGNKALH 249
```

RESULT 3
US-09-921-232-17
; Sequence 17, Application US/09921232
; Patent No. US20020102681A1
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry !
; TITLE OF INVENTION: Polypeptides Controlling Phytate
; TITLE OF INVENTION: Metabolism in Plants
; FILE REFERENCE: 0706D4
; CURRENT APPLICATION NUMBER: US/09/921,232
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 09/677,064
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Zea mays
US-09-921-232-17

Query Match 14.8%; Score 255.5; DB 10; Length 267;
Best Local Similarity 30.3%; Pred. No. 1e-16;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;

```
Qy      73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
           :: |: : ||  || :|||::|| | |   :  |     ||  |:  |: : : : | ||
Db       1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHKGQVDLVTETDKACEDLIFNHLRKHFP 60

Qy     133 THAIFGEENG---WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
```

```
Db     61 DHKFIGEEESAALGATADLTDDPTWIVDPLDGTTNFVHGFPFVCVSVGLTIGKIPTVGVV 120

Qy    190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249

Db    121 FNPIMNELFTAVRGKGAFLNGSPIKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177

Qy    250 KVPLYGCD----CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300

Db    178 NRLLYKIRSIRMCGSLALNMCGVACGRLDLCYEIGFGGPWDVAAGAVILQEAGGLVFDPS 237

Qy    301 G 301

Db    238 G 238
```

RESULT 4
US-09-921-330-17
; Sequence 17, Application US/09921330
Patent No. US20020102682A1
GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; TITLE OF INVENTION: Polypeptides Controlling Phytate
; TITLE OF INVENTION: Metabolism in Plants
; FILE REFERENCE: 0706D3
; CURRENT APPLICATION NUMBER: US/09/921,330
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 09/677,064
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Zea mays
US-09-921-330-17

Query Match 14.8%; Score 255.5; DB 10; Length 267;
Best Local Similarity 30.3%; Pred. No. 1e-16;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;

```
Qy     73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
          :: |: : ||  || :|||::|| | |   :  |     ||  |:  |: : : : | ||
Db      1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHKGQVDLVTETDKACEDLIFNHLRKHFP 60

Qy    133 THAIFGEENG---WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           |   |||        |: : |  |::||:||| :|: | |     : |     | :||:
Db     61 DHKFIGEEESAALGATADLTDDPTWIVDPLDGTTNFVHGFPFVCVSVGLTIGKIPTVGVV 120

Qy    190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
            ||: | :  | ||   |||  |   | : | :| | | :    :    |      |::
Db    121 FNPIMNELFTAVRGKGAFLNGSPIKASSQDELVKALLVTE GTNRDKTTVD---DTTNRI 177

Qy    250 KVPLYGCD----CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
           ||       | : ||    :| | :|:  | |   |:|  :   ::: ||| : |
Db    178 NRLLYKIRSIRMCGSLALNMCGVACGRLDLCYEIGFGGPWDVAAGAVILQEAGGLVFDPS 237

Qy    301 G 301
          |
Db    238 G 238
```

RESULT 5

US-09-921-329-17
; Sequence 17, Application US/09921329
; Patent No. US20020110884A1
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; TITLE OF INVENTION: Polypeptides Controlling Phytate
; TITLE OF INVENTION: Metabolism in Plants
; FILE REFERENCE: 0706D2
; CURRENT APPLICATION NUMBER: US/09/921,329
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 09/677,064
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Zea mays
US-09-921-329-17

Query Match 14.8%; Score 255.5; DB 10; Length 267;
Best Local Similarity 30.3%; Pred. No. 1e-16;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;

```
Qy     73 LATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFP 132
          :: |: : ||  || :|||::|| | |   :  |     ||  |:  |: : : : | ||
Db      1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHKGQVDLVTETDKACEDLIFNHLRKHFP 60

Qy    133 THAIFGEENG---WRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVI 189
           |   |||        |: : |  |::||:||| :|: | |     : |     | :||:
Db     61 DHKFIGEEESAALGATADLTDDPTWIVDPLDGTTNFVHGFPFVCVSVGLTIGKIPTVGVV 120

Qy    190 DQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249
            ||: | :  | ||   |||  |   | : | :| | | :    :    |      |::
Db    121 FNPIMNELFTAVRGKGAFLNGSPIKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177

Qy    250 KVPLYGCD----CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
           ||       | : ||    :| | :|:  | |   |:|  :   ::: ||| : |
Db    178 NRLLYKIRSIRMCGSLALNMCGVACGRLDLCYEIGFGGPWDVAAGAVILQEAGGLVFDPS 237

Qy    301 G 301
          |
Db    238 G 238
```

RESULT 6
US-09-987-446-2
; Sequence 2, Application US/09987446
; Patent No. US20020115162A1
; GENERAL INFORMATION:
; APPLICANT: FARWICK, Mike, et al.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES CODING FOR THE cysQ GENE
; FILE REFERENCE: 032301 WD 245
; CURRENT APPLICATION NUMBER: US/09/987,446
; CURRENT FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-987-446-2

```
  Query Match             7.3%;  Score 127;  DB 10;  Length 252;
  Best Local Similarity   28.2%;  Pred. No. 0.00013;
  Matches   37;  Conservative   21;  Mismatches   61;  Indels   12;  Gaps    4;

Qy      81 VAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGEE 140
           :  | |   ||:|:    : | ::   :       |  |:  :  |: :  |      ||
Db      10 LTHRLAQGTGEILKGV--RNVGVLRGRNLGDA--GDELAQSWIARVLEQHRPNDGFLSEE 65

Qy     141 NGWRCAEN----SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRE 196
               |:|    | | ||::||:|||| | ||:  :   |||: || |    :  | |
Db      66 ----AADNPDRLSKDRVWIIDPLDGTKEFATGRQDWAVHIALVENGVPTHAAVGLPDLGV 121

Qy     197 RWIGVDGKQTT 207
           :   | .  |
Db     122 VFHSADARAVT 132


RESULT 7
US-09-712-363-262
  Sequence 262, Application US/09712363
  Patent No. US20020164588A1
  GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
   NUMBER OF SEQ ID NOS: 292
   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 262
;   LENGTH: 1616
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-262

  Query Match             5.8%;  Score 101;  DB 9;  Length 1616;
  Best Local Similarity   24.2%;  Pred. No. 0.56;
  Matches   76;  Conservative   32;  Mismatches  116;  Indels   90;  Gaps   17;

Qy      38 VSSAVLSASG-----------RQPMSTVRASFAAGAAGR---RAAAVGELATERLVEVA 82
           |  ||| | |            | |      |  || |||   | |: |  |      ::
Db     638 VLDAVLHALGLAVEKTQASTETRLPFCWRGVSLHAGGAGRVRARFASAGADA------IS 691

Qy      83 QRAADAAG-EVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVSVILKSFPTHAIFGEEN 141
             || |  ||          |  ::  : || |   :::  : |:    |   : |  |
Db     692 VDVCDATGLPVLTVRSLVTRPITAEQLRAAVTAAGGASDQGPLEVVWS--PISVVSGGAN 749

Qy     142 G--------WR--CAENSAD---FVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGV 188
           |        |   || :  |    || |:   |  | :      |:: |  |    |:
Db     750 GSAPPAPVSWADFCAGSDGDASVVVWELESAGGQASSVV-----GSVYAATHTALEVL-- 802

Qy     189 IDQPILRERWIGVDGKQTT---------LNGQEISVRSCNLLAQAYLYTTSPHLFEADAE 239
```

```
               : |:| |   |          | |::||     || | ::  :     | ||
Db     803 -------QSWLGADRAATLVVLTHGGVGLAGEDISD-----LAAAAVWGMAR---SAQAE 847

Qy     240 DAFIRVRNKVKVPLYGCDCY--AYALLASGFVDIVVESG------LKPYDFLSLVPVIE- 290
               |  :: |   |    |  |   |   ::|  |      | |   |  :|  |
Db     848 -------NPGRIVLIDTDAAVDASVLAGVGEPQLLVRGGTVHAPRLSPAPALLALPAAES 900

Qy     291 ------GAGGSITD 298
               | ||:: |
Db     901 AWRLAAGGGGTLED 914


RESULT 8
US-09-216-393-90
; Sequence 90, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
;   LENGTH: 268
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-216-393-90

  Query Match             5.4%;  Score 93;  DB 10;  Length 268;
  Best Local Similarity   22.7%;  Pred. No. 0.24;
  Matches   42;  Conservative   32;  Mismatches   53;  Indels   58;  Gaps    8;

Qy      16 PGLASANPNPRSRLLRLRAASPVSSAVLSAS----------GRQPMSTVRASFAAGAAGR 65
           | |:  |   :| ::     :| |:  || :          |:|  |    | |  |  :
Db       7 PTLSIMN---QSTIVATSVVAPQSAVSLSRAPSRPGPSESFGKQQESRPGVSGAGLAESK 63

Qy      66 RAAAVGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAEEAMVS 125
           |  :: : : || | :::|  || |          : :: | | :  ::| :| : : :
Db      64 RVPSLTQPSLERSVTISRRKIDAVG----------MSLVPKLDRTTTSLAAKEEKFSSID 113

Qy     126 VILKSFPTHAIFGEEN------------------GW----RCAENSADFVWVLDPIDG 161
            |: | |||: |||  :                  |    |||   :          |
Db     114 KIV-SKPTHS-FGESSKLPAGIMKAKSMFPSQTLSAPWNAPARCARKDS---------FG 162

Qy     162 TKSFI 166
           ||::|
Db     163 TKAWI 167


RESULT 9
US-09-896-852-55
; Sequence 55, Application US/09896852
; Patent No. US20020025542A1
; GENERAL INFORMATION:
;  APPLICANT: Abbott Laboratories
;  APPLICANT:  Maine, Gregory T.
;  APPLICANT:  Hunt, Jeffery C.
;  APPLICANT:  Brojanac, Susan
;  APPLICANT:  Jhy-Tsing Sheu, Michael
;  APPLICANT:  Chovan, Linda E.
;  APPLICANT:  Tyner, Joan D.
;  APPLICANT:  Howard, Lawrence V.
;  TITLE OF INVENTION: ANTIGEN COCKTAILS, P35, AND USES THEREOF
;  FILE REFERENCE: 6361.US.D1
;  CURRENT APPLICATION NUMBER: US/09/896,852
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 09/086,503
;  PRIOR FILING DATE: 1998-05-28
```

```
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 55
;   LENGTH: 667
;   TYPE: PRT
;   ORGANISM: Toxoplasma gondii
US-09-896-852-55

  Query Match              5.2%;  Score 90;  DB 10;  Length 667;
  Best Local Similarity   21.8%;  Pred. No. 1.7;
  Matches   67;  Conservative   45;  Mismatches  128;  Indels   68;  Gaps   13;

Qy      12 TSPFPG--LASANPNPRSRLLRLRAASPVSSAVLSASGRQPMSTVRASFAAGAAGRRAAA 69
           ||  ||  |   |  |    :  ||    :  :: |:  : ::  ||  |||      |
Db      13 TSRLPGKPLVDINGKPMIVHVLERARESGAERIIVATDHEDVA--RAVEAAGGEVCMTRA 70

Qy      70 VGELATERLVEVAQRAADAAGEVLRKYFRQRVEIIDKEDHSPVTIADREAE---EAMVSV 126
             :  |||| || :: | :   |:       | :   |   | ||  : |:   :  | :
Db      71 DHQSGTERLAEVVEKCAFSDDTVI-------VNVQGDEPMIPATIIRQVADNLAQRQVGM 123

Qy     127 ILKSFPTH----------AIFGEENGWRCAENSADFVWVLDPI-DGTKSFITGKPLFGT 174
             : | |              :  :  |:    : |   |  |   :|  |   | | : :
Db     124 TTLAVPIHNAEEAFNPNAVKVVLDAEGYALYFSRATIPWDRDRFAEGLMSH-NGVPAYPS 182

Qy     175 L--IALLHNGKPVIGVIDQPILRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPH 232
             ::|  ||:|            |  |: |                      ::| :  ||
Db     183 YAQVSLSSNGEP----------RHRGIRG------------------SFLMSVKPH 210

Qy     233 LFEAD--AEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIE 290
             |  ::| :  : : |: |: | |     : | |   :| |         ::|    |
Db     211 ANADDFASDDNYEPLPSFVEAPVRGPD----QVPARGEAALVTEETPAQQPA LGSA-E 265

Qy     291 GAGGSITD 298
           | | | |:
Db     266 GEGTSTTE 273


RESULT 10
US-09-925-301-1115
; Sequence 1115, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1115
;   LENGTH: 777
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (9)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (14)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (21)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (32)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-301-1115

  Query Match              5.2%;  Score 90;  DB 10;  Length 777;
  Best Local Similarity   20.8%;  Pred. No. 2.1;
  Matches   60;  Conservative   39;  Mismatches   92;  Indels   98;  Gaps   12;

Qy      33 RAASPVSSA-------VLSASGRQPMSTVRASFAAGAAGRRAAAVGELATERLVEVAQRA 85
           ||  |: |:       |||:  | ||   :|| :|  |     :|    ::     ::
Db      30 RAXPPLGSSPLGRRFRVLSSLRRSPMFEEKASSPSGKMGGEEKPIGAGEEKQKEGGKKKN 89

Qy      86 ADAAGEVLR-------KYFRQRVEI--IDKEDHSPVTIADREAEEAMVSVIL-------- 128
             : :|:  |        :|   |:|:  | | :|  :     | :   : | |
Db      90 KEGSGDGGRAELNPWPFYIYTRLEMYNILKAEHDSILAEKAEKDSKPIKVTLPDGKQVDA 149

Qy     129 ---KSFPTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPV 185
              |: |     |   |   |:|                          |:|| ::|   |
Db     150 ESWKTTPYQIACGISQG--LADN-----------------------TVIAKVNN---V 179

Qy     186 IGVIDQPILRERWIGVDGKQTTLNGQEISVRSCNL--------LAQAYLYTTSPHLFEAD 237
             :  :|:|:  :                        | |          |||  : :| |:
Db     180 VWDLDRPLEED--------------------CTLELLKFEDEEAQAVYWHSSAHIM--- 215

Qy     238 AEDAFIRVRNKVKVPLYGCDCYAYALLASGFVDIVV-ESGLKPYDFLSL 285
             :|  ||          || ||   :    : |: : | |:   || ||
Db     216 -GEAMERVYG-------GCLCYGPPIENGFYYDMYLEEGGVSSNDFSSL 256


RESULT 11
US-10-027-806-4
; Sequence 4, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
;  APPLICANT: Swanson, Ronald V.
;  APPLICANT:  Feldman, Robert A.
;  APPLICANT:  Schleper, Christa
;  TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
;  FILE REFERENCE: DCORP.002A
;  CURRENT APPLICATION NUMBER: US/10/027,806
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
;  PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
;  NUMBER OF SEQ ID NOS: 123
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 3472
;   TYPE: PRT
;   ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

  Query Match              5.1%;  Score 88;  DB 9;  Length 3472;
  Best Local Similarity   25.5%;  Pred. No. 28;
  Matches   60;  Conservative   24;  Mismatches   81;  Indels   70;  Gaps   12;

Qy     116 DREAEEAMVSVILKSF---PTHAIFGEENGWRCAENSADFVWVLDPIDGTKSFITGKPLF 172
           | |  |  :|  | :    ||  :| :|| :      | ||         : | | :|
Db    2435 DTENAEHFISTDLLTADRGPTGLVFSDENDFFSTGARAQFV---------RQFTTNRPYD 2485

Qy     173 GTLIALLHNGKPVIGVIDQPILRERWIGV----DGKQTTLNGQEISVRSCNLLAQAYLYT 228
             : | |  ||   : |   |       |:    || :  ::|||          |  | |:
Db    2486 ASTITLSDNGLYKVSVDGLP------SGIRFTPDGMKMFISGQE--------TAMIYQYS 2531

Qy     229 -TSPHLFEADAEDAFIRVRNKVKVPLYGCDCYAYALLASGFVD---------------IV 272
            ||:       |    ||::|::       :  | |: |  :                 |
Db    2532 LPSPY-------DTSGAVRDRVEIV---AGLFRNAGLSVGLNEPSPSGFDFSEDGMELYV 2581

Qy     273 VESGLKPYDFLSLVPVIEGA--GGSITDWRGDKLHWPVTAESRPTSFNVVAAGDA 325
           |||    ||     :| |  |||   :|         || |    ||  |||
Db    2582 TGSGLVHRYFLPSPYGLEDAAYGGSFHTFR----------ESTP--LGVVVRGDA 2624


RESULT 12
US-09-817-464-2
; Sequence 2, Application US/09817464
```

```
; Patent No. US20020127638A1
; GENERAL INFORMATION:
;  APPLICANT: Flor, Peter J.
;  APPLICANT: Kuhn, Ranier
;  APPLICANT: Lindaur, Kristen
;  APPLICANT: Puttner, Irene
;  APPLICANT: Knopfel, Thomas
;  TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
;  TITLE OF INVENTION: HMR6, HMR7) and Related DNA Compounds
;  FILE REFERENCE: 4-19679/A/PCT
;  CURRENT APPLICATION NUMBER: US/09/817,464
;  CURRENT FILING DATE: 2001-03-26
;  EARLIER APPLICATION NUMBER: US/08/617,785
;  EARLIER FILING DATE: 1996-03-19
;  EARLIER APPLICATION NUMBER: EPO 9416553.7
;  EARLIER FILING DATE: 1994-08-19
;  EARLIER APPLICATION NUMBER: EPO 93810663.0
;  EARLIER FILING DATE: 1993-09-20
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
   SEQ ID NO 2
     LENGTH: 912
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-09-817-464-2

  Query Match             5.0%;  Score 86;  DB 10;  Length 912;
  Best Local Similarity   24.0%;  Pred. No. 6.4;
  Matches   65;  Conservative   38;  Mismatches   84;  Indels   84;  Gaps   17;

Qy      40 SAVLSASGRQPMSTVRASFAAGAAGRRAAAVGELATERLVEVAQRAADA--AGEVLRKYF 97
           |::    |   |:::|              |::    |  |  | ::    ||: ::
Db      29 SSLGKPKGHPHMNSIRID------------GDITLGGLFPVHGRGSEGKPCGELKKEKG 75

Qy      98 RQRVE----IIDKEDHSP-----VTIADREAEEAMVSVILKSFPTHAIFGEENGWRCAEN 148
             |:|     :|: :: |     :|:  |  :          |  |||:           |
Db      76 IHRLEAMLFALDRINNDPDLLPNITLGARILDTC-------SRDTHAL----------EQ 118

Qy     149 SADFVWVLDPIDGTKSFITGKPLFGTLIALLHNGKPVIGVIDQPILRERWIGVDGKQTTL 208
           |  ||  |   |||:            :     | |   :| :|   || :|| |
Db     119 SLTFVQALIEKDGTE-----------VRCGSGGPP---IITKP---ERVVGVIG----A 156

Qy     209 NGQEISVRSCNLL-----AQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLYGCDCYAYAL 263
           :|  :|:   |:|       |    :|:| | :    | | ||     ||    | |
Db     157 SGSSVSIMVANILRLFKIPQISYASTAPDLSDNSRYDFFSRV-----VP---SDTYQ--- 205

Qy     264 LASGFVDIVVESGLKPYDFLSLVPVIEGAGG 294
            |   ||||    || ::::| |   ||: |
Db     206 -AQAMVDIV--RALK-WNYVSTV-ASEGSYG 231


RESULT 13
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-33

  Query Match             4.9%;  Score 84;  DB 9;  Length 3739;
  Best Local Similarity   32.3%;  Pred. No. 75;
  Matches   31;  Conservative   15;  Mismatches   40;  Indels   10;  Gaps    4;

Qy      46 SGRQPMSTVRASFAAGAAGRRAA--AVGELATERLVEVAQRAADA--AGEVLRKYFRQRV 101
           ||  |  ||  :   || |   |   |||    ||: |::|  ||  |||:: :    :
Db    3271 SGLDPEGTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHE-----L 3325

Qy     102 EIIDKE-DHSPVTIADREAEEAMVSVILKSFPTHAI 136
           | :  :   :  :|||||  |::  |    |  |:
Db    3326 EALGADVSVAACDVADREALTAVLDSIPAEHPLTAV 3361


RESULT 14
US-09-861-289-33
; Sequence 33, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-33

  Query Match             4.9%;  Score 84;  DB 10;  Length 3739;
  Best Local Similarity   32.3%;  Pred. No. 75;
  Matches   31;  Conservative   15;  Mismatches   40;  Indels   10;  Gaps    4;

Qy      46 SGRQPMSTVRASFAAGAAGRRAA--AVGELATERLVEVAQRAADA--AGEVLRKYFRQRV 101
           ||  |  ||  :   || |   |   |||    ||: |::|  ||  |||:: :    :
Db    3271 SGLDPEGTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHE-----L 3325

Qy     102 EIIDKE-DHSPVTIADREAEEAMVSVILKSFPTHAI 136
           | :  :   :  :|||||  |::  |    |  |:
Db    3326 EALGADVSVAACDVADREALTAVLDSIPAEHPLTAV 3361


RESULT 15
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-6
```

```
  Query Match             4.9%;  Score 84;  DB 9;  Length 11877;
  Best Local Similarity   32.3%;  Pred. No. 4e+02;
  Matches   31;  Conservative   15;  Mismatches   40;  Indels   10;  Gaps    4;

Qy      46 SGRQPMSTVRASFAAGAAGRRAA--AVGELATERLVEVAQRAADA--AGEVLRKYFRQRV 101
           ||  |  ||  :   || |   |   |||    ||: |::|  ||  |||:: :    :
Db    8220 SGLDPEGTVLLTGGTGALGGIVARHVVGEWGVRRLLLVSRRGTDAPGAGELVHE-----L 8274

Qy     102 EIIDKE-DHSPVTIADREAEEAMVSVILKSFPTHAI 136
           | :  :   :   :|||||  |::  |    |  |:
Db    8275 EALGADVSVAACDVADREALTAVLDSIPAEHPLTAV 8310


Search completed: November 17, 2002, 16:44:49
Job time : 56 secs
```